Structural Characterisation of members of the DHDPS/NAL subfamily of proteins

A thesis submitted in partial fulfilment

of the requirements for the degree of

Master of Science

at the

School of Biological Sciences,

University of Canterbury

By

Cameron MacDonald

UNIVERSITY OF CANTERBURY

Te Whare Wānanga o Waitaha

CHRISTCHURCH NEW ZEALAND

2017

Table of contents

Acknowledgements

First, I would like to thank both of my supervisors, Dr Grant Pearce and Assoc. Prof. Renwick Dobson for helping my through this year. Thank you for guiding through this year, both in the lab and out of it. Your advice and guidance helped me get to where I am today.

My thanks to everyone in the Dobson lab for helping to pass the time talking about some stuff and the things as well as the occasional supervised events at the lab meeting.

A thanks to everyone in room 630: William Finnis, James Davies, Serena Watkin, Hamish Cleland and David Coombes for making this year a lot less stressful. Thank you for helping me with any questions I had. The various things we talked about helped make the past year quite enjoyable. Finally, a thanks to my family and friends for all the support over the past year.

Thank you all.

Abstract

The DHDPS/NAL subfamily of enzymes share a very similar structure while catalysing a wide variety of reactions in different biochemical pathways. The structural similarity between members allows us to understand how this subfamily of enzymes has evolved. This project characterises some of the more obscure enzymes in the DHDPS/NAL subfamily in order to investigate evolutionary relationships within this subfamily of proteins. This will be done though techniques to investigate the structural states of these proteins as well as their solution based structure.

There are two protein groups under investigation: trans-hydroxybenzylidenepyruvate hydratase-aldolase (tHBP-HA), 1-pyrroline-4-hydroxy-2-carboxylate deaminase (HypD). tHBPHA catalyses the final step of the naphthalene degradation pathway, it has been previously characterised as a trimer. Analytical ultracentrifugation (AUC) has been used to probe the oligomeric states of the proteins to confirm if they are similar to the tetrameric DHDPS/NAL. tHBPHA was found to be consistent with a tetrameric protein. This was further shown through use of Small angle x-ray scattering (SAXS). Structure for tHBP-HA from *P. fluorescens* was solved through X-ray crystallography to 2.2Â, allowing for investigation of the active site. The replacement of a tyrosine from the adjacent monomer was the only difference in the conserved amino acids within the active site. It was found that tHBPHA does not show activity towards the DHDPS substrates.

The second protein investigated is HypD. HypD is involved in L-Hydroxyproline metabolism in bacteria and different groups have labelled it as either a tetrameric or hexameric protein. AUC showed that the four HypD proteins were consistent with hexameric proteins and were largely different to the reference protein *E. coli* DHDPS. This was again further shown using SAXS. The crystal structure for *M. lupine* HypD was produced to 2.9Å. The protein fits the hexameric architype, further showing that HypD are hexameric. The active site showed no interaction with the opposing monomer in the dimer interface, a contrast to the other members of the DHDPS/NAL subfamily. This shows that HypD is a novel member of the DHDPS/NAL subfamily. It was found that HypD does not show activity towards the DHDPS substrates.

Abbreviations

ºC	Degrees Celsius
1NAL	*E. coli* NAL
1YXC	*E. coli* DHDPS
2HMC	*A. tumefaciens* HypD
4MPQ	*B. melitensis* HypD
5CZJ	*S. meliloti* HypD
A.U.	Absorbance units
AUC	Analytical Ultracentrifugation
c(M)	Continuous mass distribution
c(s)	Continuous sedimentation distribution
$CaCl_2$	Calcium chloride
Cryo	Cryoprotectant
DHDPR	Dihydrodipicolinate reductase
DHDPS	Dihydrodipicolinate synthase
DOGDH	D-4-deoxy-5-oxoglucarate dehydratase
DSF	Differential scanning fluorimetry
His-tag	Histidine tagged protein
HypD	Δ^1-Pyrroline-4-hydroxy-2-carboxylate Deaminase
I_0	Zero scattering angle
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	Kilo Dalton
K_m	Michaelis Constant
LB	Luria-Bertani broth

m9zb?	High nutrient growth media
MES	2-ethanesulfonic acid
mPa	Mega pascal
Na_2HPO_4	Sodium phosphate
NaCl	Sodium Chloride
NAL	*N-acetylneuraminate lyase*
nm	Nanometre
OD_{600}	Absorbance at 600 nanometres
p(r)	Pair distribution function plot
pdb	Protein data base
R_g	Radius of gyration
rpm	Round per minute
SAXS	Small Angle x-ray Scattering
SEC	Size-exclusion chromatography
Sypro orange	SYPRO Orange Protein Gel Stain
TEV	TEV protease
tHBPHA	Trans-Hydroxybenzylidenepyruvate hydratase-aldolase
T_m	Melting temperature
tris-HCl	Trizma hydrochloride
UV	Ultraviolet
v_{max}	Maximal Velocity

Table of Figures

List of tables

1. Chapter one : introduction

1.1. A brief overview of the thesis

Study of the dihydrodipicolinate synthase (DHDPS) / N-acetylneuraminate lyase (NAL) subfamily of proteins provides an effective way to investigate protein evolution between members of the same proteins subfamily. This is due to the high sequence and structural similarities between members of this sub family of TIM barrel proteins. While DHDPS and NAL proteins have been well characterised due to their importance in different metabolic pathways such as bacteria, archaea and plants, other members of this subfamily haven't been as thoroughly characterised. Due to the lack of information on these proteins, their structural and biophysical characteristics are relatively unknown and contradictions between different studies appear.

The aim of this master's thesis is to investigate some of the more obscure members of the DHDPS/NAL subfamily of proteins in order to investigate structural similarities between these members. These members are: Trans-o-hydroxybenzylidenepyruvate hydratase-aldolase (tHBPHA) and Δ1-pyrroline-4-hydroxy-2-carboxylate deaminase (HypD). These proteins have not been as thoroughly studied as DHDPS or NAL, because of this the structural properties of these proteins are inconsistent.

1.2. The TIM barrel fold

A TIM barrel fold is a conserved protein fold in which eight α-helices surround eight parallel β-sheets which form a cavity through the protein. The active site of these proteins is generally located within this barrel. This protein fold contains a wide variety of different protein subfamilies which catalyse a wide variety of reactions. The high number of proteins which exhibit this fold is estimated to be about 10% of all currently known enzymes (Copley and Bork 2000, Nagano, Orengo et al. 2002). All TIM barrel proteins have been suggested to have evolved from a common ancestor. This explains why such a large number of proteins share a highly conserved fold. The fold itself has likely evolved

into its current form and it has been theorised that the barrel evolved from its ancestors who possibly possessed half barrel (Lang, Thoma et al. 2000, Hocker, Beismann-Driemeyer et al. 2001).

One subfamily of proteins within this 10% is the DHDPS / NAL subfamily of proteins. This is a family which (despite originating from a highly diverse background in numerous metabolic pathways) all exhibit the conserved TIM barrel fold and a similar reaction chemistry. There are numerous members of this family which catalyse a wide variety of reactions. Despite catalysing a variety of different reactions, they exhibit structural and biochemical similarities (Nagano, Orengo et al. 2002). A majority of these proteins are classed as aldolase enzymes. Examples of the different reactions catalysed by members of this sub family is the reaction of DHDPS which catalyses the condensation of pyruvate with l-aspartate β-semialdehyde into the product Dihydrodipicolinate (Blickling, Renner et al. 1997). Another example is the reaction of NAL which catalyses the aldol cleavage of sialic acid (N-acetylneuraminate) to N-acetyl-D-mannosamine + pyruvate(Uchida, Tsukada et al. 1984). The member tHBPHA catalyses the final step of the naphthalene degradation pathway(Eaton and Chapman 1992). HypD is involved with the bacterial degradation of hydroxyl proline (Watanabe, Kodaki et al. 2006, Watanabe, Yamada et al. 2007) while yet another member (HOG aldolase) catalyses the final step of L-hydroxyproline metabolism in mammals, yielding Pyruvate and glyoylate as products (Riedel, Johnson et al. 2011). All of these proteins share structural similarities whilst catalysing different reactions, showing the diversity of members of this subfamily.

1.2.1. The DHDPS / NAL subfamily of proteins

These members can be described as class one aldolase enzymes. There are two different classes of Aldolase enzymes: class one and class two. Class one Aldolase enzymes initially activate a donor substrate through formation of a Schiff base intermediate via linkage of an active site residue (eg lysine) to a substrate, followed by condensation or cleavage of a substrate. The basic aldol reaction involves the condensation of an aldehyde with a ketone (a reaction with an aldol receiver and an aldol donor) resulting in the formation of a new carbon to carbon bond (Bolt, Berry et al. 2008). Class one Aldolase enzymes do not require the presence of a bivalent metal cofactor which are required for activity in class two aldolase (Gefflaut, Blonski et al. 1995).

1.2.1.1. Dihydrodipicolinate synthase

One of the well characterised members of this subfamily is DHDPS which is involved in the biosynthesis of lysine. Lysine is a key amino acid used for the production of cellular proteins that is not synthesised by the human body and therefore must be obtained from dietary sources(Viola 2001). Lysine is a nutritionally limiting factor in cereal crops, thus (though controversial) genetically modifying these crops in order to boost lysine production of these crops is beneficial in increasing the amount of dietary lysine available. (Bright and Shewry 1983, Viola 2001). The diaminopimelate pathway (DAP) is one of two lysine biosynthetic pathways which exists in bacteria, plants and archaea which is involved in the synthesis of lysine (Velasco, Leguina et al. 2002). DHDPS is the first committed step of this reaction (figure 1). As previously mentioned, DHDPS catalyses the condensation reaction between pyruvate with l-aspartate β-semialdehyde (ASA), forming the product (2S,4S)-4-hydroxy-2,3,4,5-tetrahydrodipicolinate (Blickling, Renner et al. 1997). As with other members of the Type one Aldolase subfamily, the reaction involves the formation of a Schiff base intermediate between the catalytically active lysine within the active site and the substrate pyruvate (Blickling, Renner et al. 1997).

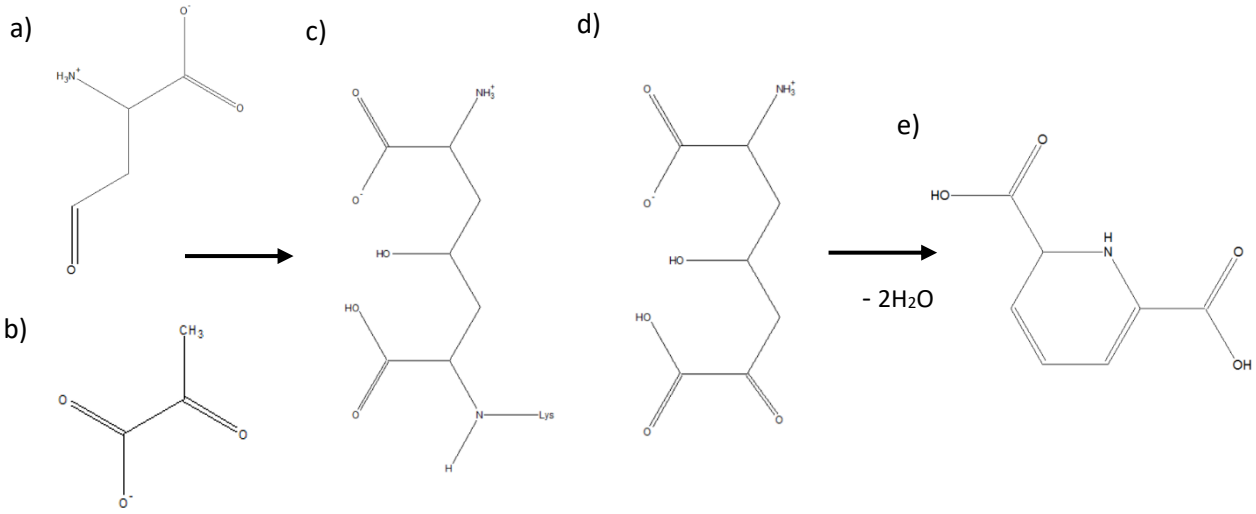


Figure 1

The catalytic reaction of DHDPS is shown, beginning with the two substrates l-aspartate β-semialdehyde (a) and pyruvate (b). Lysine within the active site binds to the pyruvate (c) and a aldol bond formation takes place (d). The molecule is dehydrated and the product (2S,4S)-4-hydroxy-2,3,4,5-tetrahydrodipicolinate is formed (e).

Following this, ASA then binds, is dehydrated and the cyclic (2S,4S)-4-hydroxy-2,3,4,5-tetrahydrodipicolinate is formed along with a water molecule. The lysine residue is important for catalytic activity, but its removal won't prevent pyruvate binding in the active site (da Costa, Muscroft-Taylor et al. 2010). An active site mutation which targeted and removed the highly conserved lysine residual only resulted in a significant decrease in activity but was still able to be catalytically active (da Costa, Muscroft-Taylor et al. 2010).

1.2.1.2. N-acetylneuraminate lyase

Another member of this subfamily is N-acetylneuraminate lyase (NAL). NAL catalyses the reverse aldol reaction where sialic acid is cleaved into N -acetylmannosamine and pyruvate (figure 2) (Izard, Lawrence et al. 1994, Lawrence, Barbosa et al. 1997). Sialic acid are nine-carbon amino sugars often found on the end of eukaryotic cell surface glycoconjugates (Vimr, Kalivoda et al. 2004) They are present in abundance within the human gastrointestinal tract and respiratory system(Jeong, Oh et al. 2009), locations where the presence of glucose is a limiting factor for pathogenic colonisation. The lack of glucose is combated through evolutionary development of pathways in which to degrade sialic acid for use as a nutrient source (Vimr, Kalivoda et al. 2004, Almagro-Moreno and Boyd 2009). Sialic acid is imported into the bacterial cell through sialic acid importers with NAL providing the first step in sialic acid degradation following its import into the cell. This reaction is reversible (like the other reactions from this subfamily) where the reverse reaction synthesises sialic acid. The NAL active site shares common elements to DHDPS such as the highly conserved lysine and tyrosine residuals with the only active site residual not conserved being serine (in place of threonine in DHDPS).

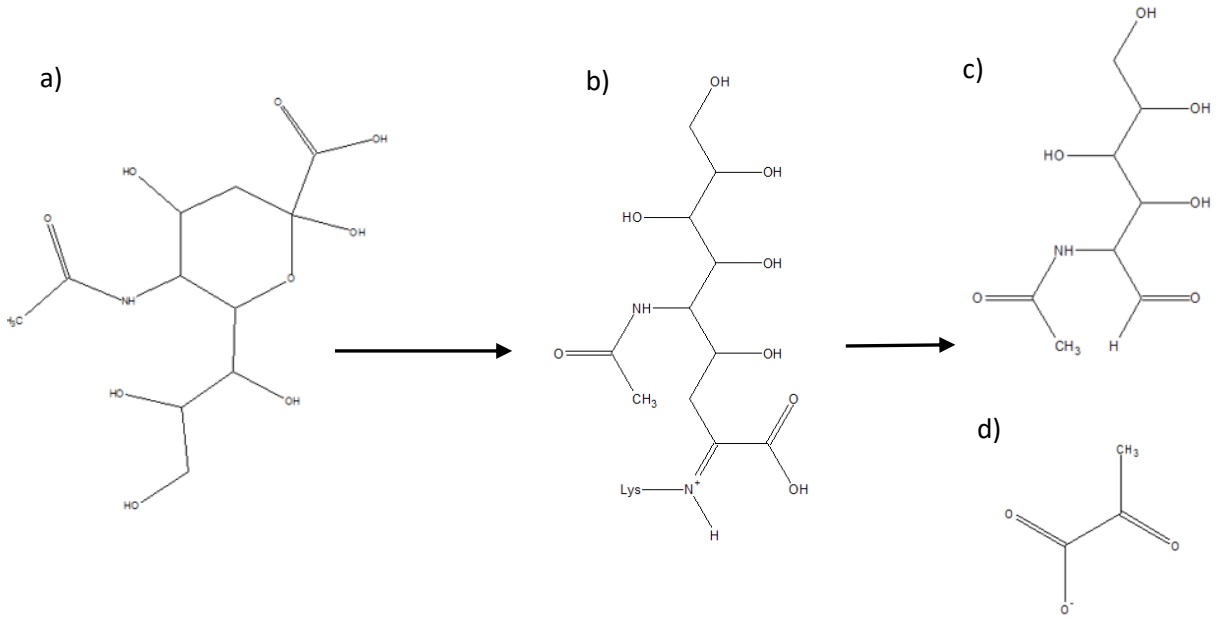


Figure 2

The catalytic reaction of NAL is shown, Sialic acid (a) has its ring broken and is then and bound to the active site lysine (b) via Schiff base pairing to the terminal ε-amine. This is then cleaved into N-acetylmannosamine (c) and pyruvate (d)

1.2.1.3. Structural similarities between *E. coli* DHDPS and NAL

All members of the DHDPS/NAL subfamily exhibit the same TIM barrel fold. This fold in highly conserved throughout members of this subfamily along with certain key amino acids present within the active site. All members of this subfamily all show the presence of lysine within the active site. Lysine is key for the aldolase activity of members of this subfamily. This is due to its key part in binding the binding of the substrate (primarily pyruvate in most members) and formation of a Schiff base between lysine residual and substrate(Laber, Gomisruth et al. 1992). The importance of this amino acid is shown through mutagenesis studies.

In *E. coli* NAL, when the conserved lysine residual was mutated to alanine, the enzymatic activity was rendered completely inactive. Another mutation to arginine resulted in the enzyme retaining about 3% of its original activity towards sialic acid (Kruger, Schauer et al. 2001). This shows a similar case for *E. coli* DHDPS which in which the same mutations were performed. There was significant decrease in activity for both mutations lysine into alanine ($0.06 \pm 0.02 s^{-1}$) and arginine ($0.16 \pm 0.06 s^{-1}$) when compared to the original activity of *E. coli* DHDPS ($45 \pm 3 s^{-1}$) (da Costa, Muscroft-Taylor et al. 2010). These studies show the importance of lysine within the active site of members of this subfamily.

Both *E. coli* DHDPS and NAL share a high sequence and structural similarity. Both are tetrameric proteins and follow the same catalytic mechanism (Laber, Gomisruth et al. 1992, Izard, Lawrence et al. 1994). Figure 3, a) Shows both *E. coli* DHDPS (1YXC) and NAL (1NAL) monomers overlaid showing the high similarity with both TIM barrel proteins. The active site consists of three key amino acids and a tyrosine which protrudes into the active site from the monomer–monomer interface. Figure 3, b) shows the alignment of both active sites. The key residues in *E. coli* DHDPS are: T44, Y133, K161 from within one monomer and Y107 from the opposing monomer (Laber, Gomisruth et al. 1992). *E. coli* NAL show the key enzymatic residues of: S47, Y137, and K165 from within the monomer and Y110 from the opposing monomer (Izard, Lawrence et al. 1994). The highly conserved active site residues shared between the two proteins is shown with the position of these key amino acids. Because of this high similarity between members of this subfamily, it provides a good platform to examine enzymatic evolution within a subfamily of proteins.

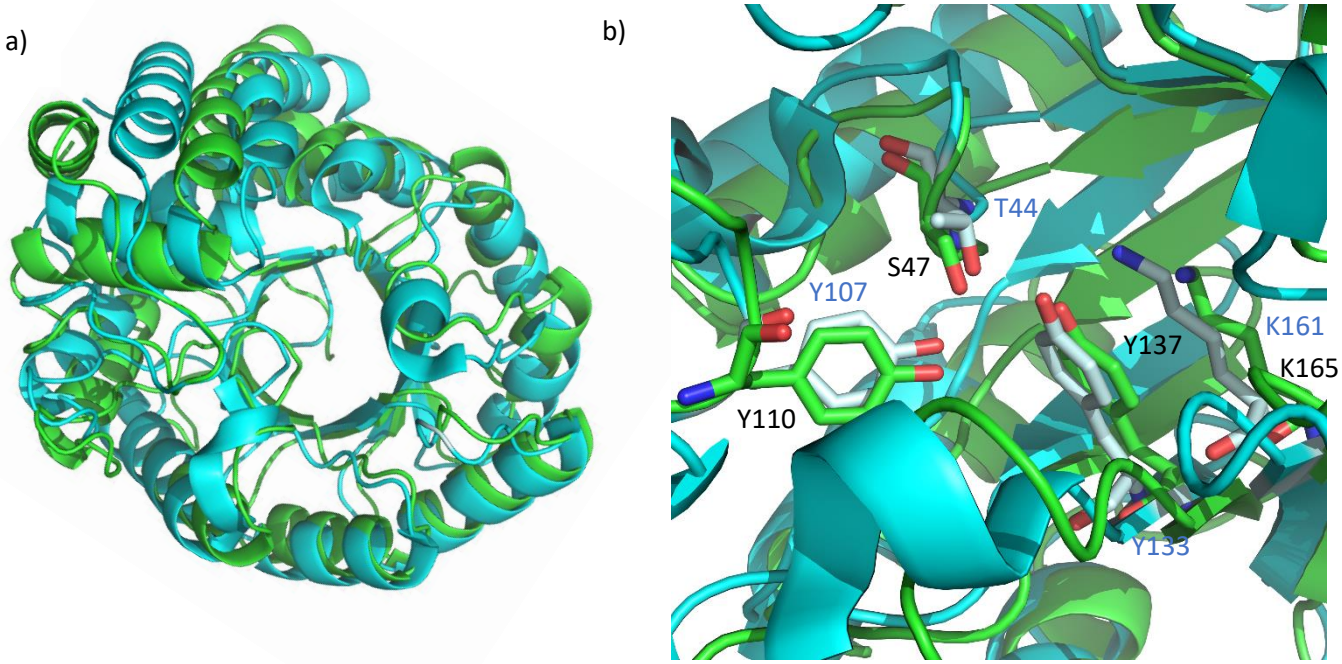


Figure 3

E. coli DHDPS (1yxc: cyan) and NAL (1NAL: green) monomers were aligned (a) to show the high similarity within the conserved TIM barrel fold. The active sites of the two proteins were aligned (b) showing the highly conserved and similar amino acids within. DHDPS residues are coloured green (black text) while NAL residues are coloured light cyan (blue text).

The similarities between the members of this subfamily is further shown with figure 4. This figure shows four different members of this subfamily. *E. coli* DHDPS (a) and *A. thaliana* DHDPS (b) show a different orientation. There is some variation between these members such as DHDPS from *A. thaliana (*plant) and *E. coli* (bacteria) which show a flipped orientation (Griffin, Billakanti et al. 2012). As previously mentioned, E. coli NAL (c) shares a similar orientation to *E. coli* DHDPS which is shown with a comparison between the two tetrameric structures. *S. meliloti* MOSA (d) is another member of this subfamily and is present in a similar tetrameric fold. They all are members of the DHDPS/NAL subfamily, exhibiting a homotetrameric orientation. The highly conserved fold is prevalent within all members of this subfamily.

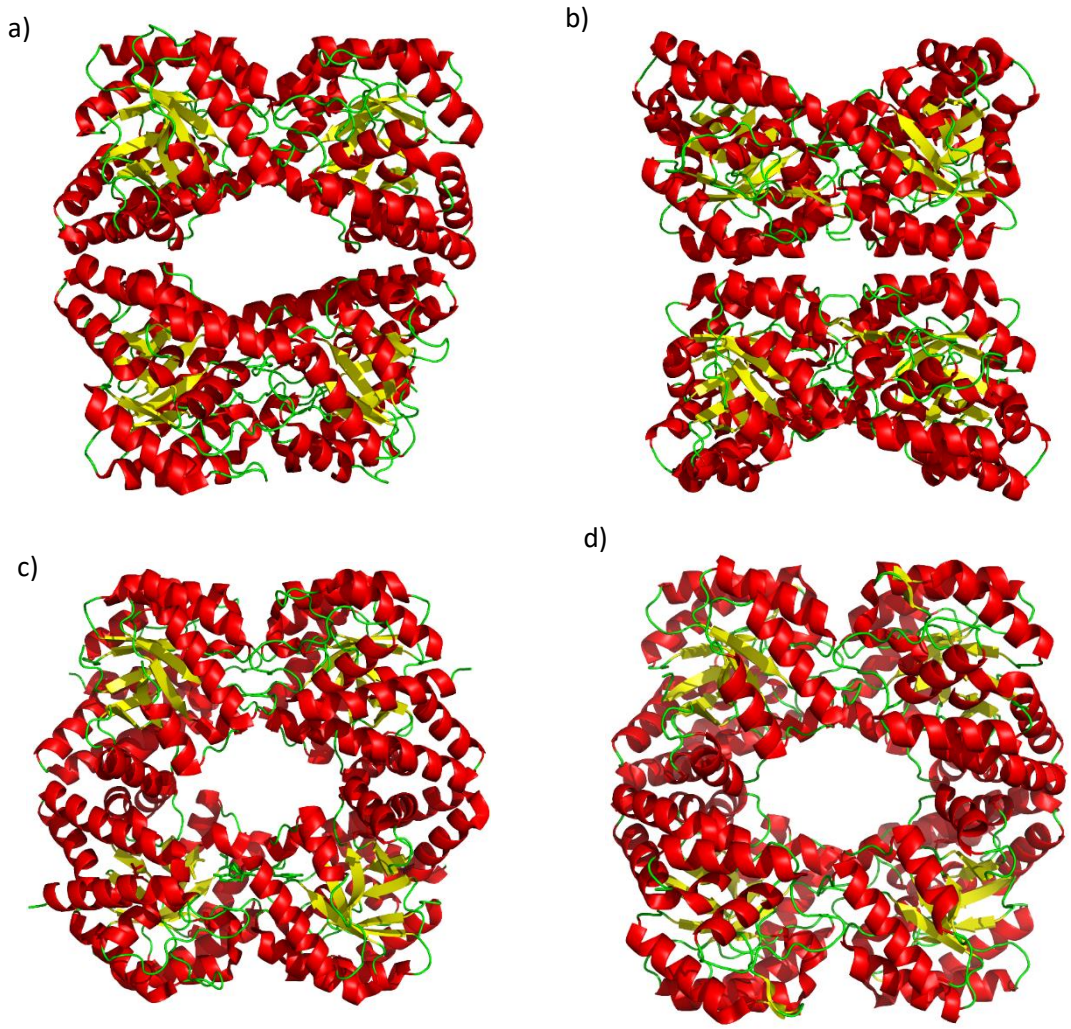


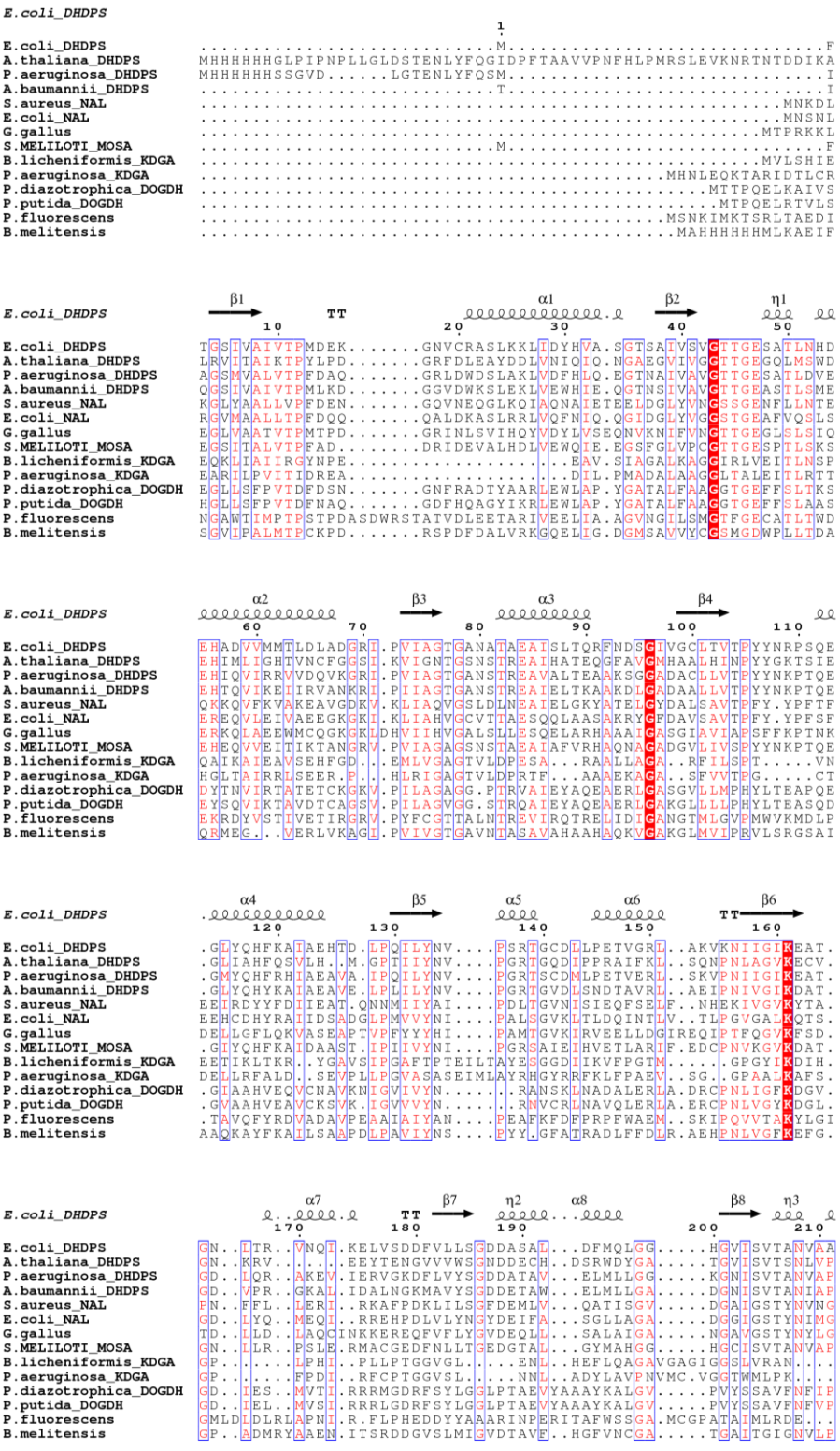
Figure 4

This figure shows the tetrameric structures of different members of the DHDPS/NAL subfamily of proteins. a) *E. coli* DHDPS, b) *A. thaliana* DHDPS, c) *E. coli NAL* and *S. meliloti* MOSA. α-helices are coloured in red while β-sheets are coloured yellow. All four proteins show a similar oligomeric state.

1.2.1.4. Sequence alignment of members of the DHDPS/NAL subfamily.

To further highlight the similarities between the different members of this subfamily, a sequence alignment of a few members of this subfamily was performed. Figure 5 shows a sequence alignment of a number of these subfamily members. This figure shows conserved residues which are highlighted in red surrounded by a blue box. The secondary structure of *E. coli* DHDHPS is also shown along the top of the sequence alignment to represent where the different β-sheets and α-helix are positioned. There are a large number of residuals which are conserved throughout the subfamily.

In *E. coli* DHDPS the key active site residuals are: T44, Y133, K161 and Y107, they are highly conserved throughout all DHDPS and some other members. Y133 is conserved throughout most members, being present within the active site. Y107 from the opposite monomer is also conserved throughout a number of members but with some variation. K161 is a key residual that forms a covalent intermediate with pyruvate (Laber, Gomisruth et al. 1992, Blickling, Renner et al. 1997) which is why it is conserved throughout all members. Its absence results in a significant drop in enzymatic activity as was mentioned previously (da Costa, Muscroft-Taylor et al. 2010). These conserved active site residues is expected due to all members of the DHDPS/NAL subfamily react to create/cleave an aldol (Nagano, Orengo et al. 2002).


E.coli_DHDPS
A.thaliana_DHDPS
P.aeruginosa_DHDPS
A.baumannii_DHDPS
S.aureus_NAL
E.coli_NAL
G.gallus
S.MELILOTI_MOSA
B.licheniformis_KDGA
P.aeruginosa_KDGA
P.diazotrophica_DOGDH
P.putida_DOGDH
P.fluorescens
B.melitensis

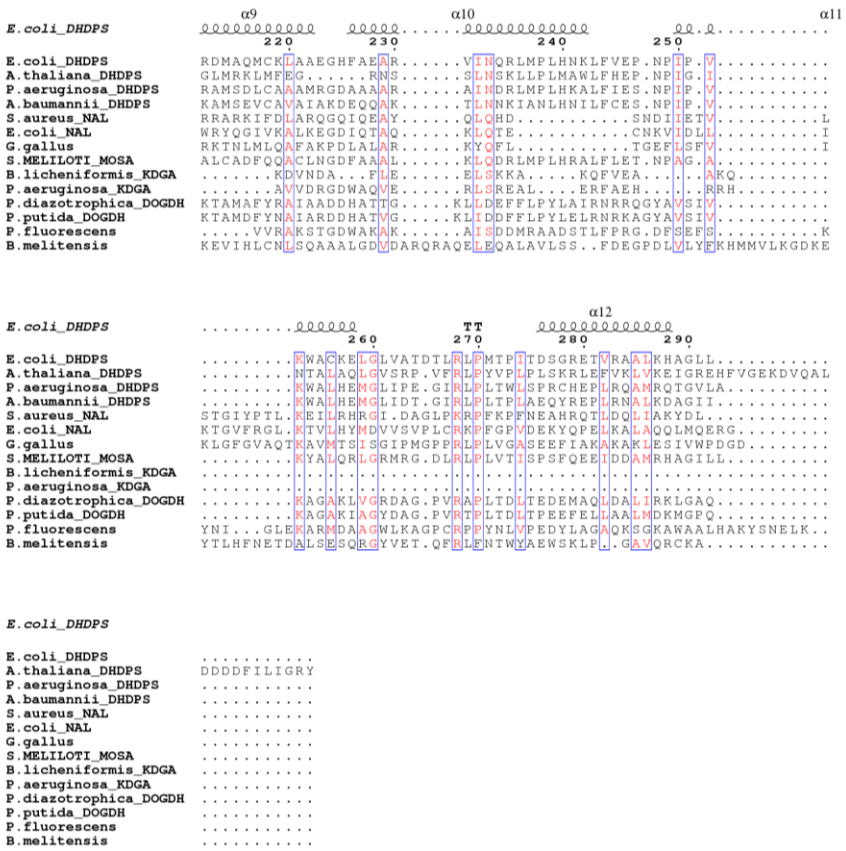


Figure 5

This figure shows the sequence alignments of members of the DHDPS/NAL subfamily of proteins. In descending order*; E. coli* DHDPS*, A. thaliana* DHDPS*, Pseudomonas aeruginosa* DHDPS*, Acinetobacter baumannii* DHDPS*, Staphylococcus aureus* NAL*, E. coli* NAL*, Xenopus tropicalis* NAL*, Sinorhizobium meliloti* MOSA*, Bacillus licheniformis* KDGA*, Pseudomonas aeruginosa* KDGA*, Pleomorphomonas diazotrophica* DOGDH *and Pseudomonas putida* DOGDH*.* This information is displayed using espript software (Robert and Gouet 2014) with secondary structure of *E. coli* DHDPS being shown along the top. Highly conserved residues throughout most proteins are shown in red surrounded by a blue box while residuals conserved throughout all members are highlighted by a red background.

1.3. Aim of thesis

The aim of this thesis is to investigate some of the more obscure members of this subfamily in order to determine their structural characteristics in relation to this subfamily. This will involve use of various techniques to probe the structural characteristics such as: Analytical ultracentrifugation (AUC), small angle x-ray scattering (SAXS) and x-ray crystallography. The reason for the structural characterisation is due to different studies reporting interesting or contrasting facts about these proteins that diverge from the typical members of the DHDPS/NAL subfamily. The possibility for DHDPS activity will also be determined through kinetic assay and DSF.

1.4. Proteins of interest

1.4.1. trans-Hydroxybenzylidenepyruvate hydratase-aldolase (tHBPHA)

The first protein family which will be investigated is trans-Hydroxybenzylidenepyruvate hydratase-aldolase (tHBPHA). This protein catalyses the final step of the naphthalene degradation pathway and is a member of the DHDPS/NAL subfamily of proteins therefor expressing the conserved fold and a similar active site architecture. Previous studies have characterised some aspects of this protein but have produced some interesting results. (Chen et al) investigated tHBPHA from Pseudomonas uesicularis and estimated that the total molecular weight of tHBPHA was about 120 kDa with a likely trimeric orientation present. This estimate was based off Gel results but, if correct, could result in a trimeric member of the DHDPDS/NAL subfamily. How this oligomeric state will affect the catalytic mechanism is interesting due to how a normal tetramer is aligned in this subfamily.

1.4.2. Δ^1-pyrroline-4-hydroxy-2-carboxylate Deaminase (HypD)

The protein Δ^1-pyrroline-4-hydroxy-2-carboxylate Deaminase (HypD) is another member of the DHDPS/NAL subfamily. It is involved in the degradation of hydroxyproline. This protein seems to have evolved from the same subfamily of DHDPS/NAL but its catalytic mechanism is different in the fact that pyruvate is not present as a product or substrate (Watanabe, Kodaki et al. 2006, Watanabe, Yamada et al. 2007). One interesting point about HypD proteins which have been unclear is the oligomeric state. Two different papers have produced contrasting evidence on which oligomeric state HypD exists as. Chen et al investigated HypD from *S.meliloti* and concluded that the oligomeric state is in a tetrameric orientation similar to the other members of the DHDPS/NAL subfamily (Chen, White et al. 2016). This contrasts to *A.tumefaciens* HypD which is reported to exist in a hexameric orientation (pdb:2HMC). This result contrasts to other members of this subfamily and is another one of the main points this thesis will investigate. The possible mechanism of HypD is also interesting due to only lysine and tyrosine being conserved within the active site, suggesting that different amino acids are involved in catalytic activity.

2. Chapter two: tHBPHA is a trimeric protein?

2.1. Introduction into the second chapter.

Due to the high structural similarity shared between the members of the DHDPS/NAL subfamily of proteins, analysing the evolution of proteins in this subfamily is possible. Some of the more obscure members of this subfamily exhibit possibly interesting characteristics which is not present with the other, well characterised members of this sub family. This chapter aims to structurally characterise one possibly novel member of this subfamily which has been reported as a possible trimeric protein.

One example of this is trans-hydroxybenzylidenepyruvate hydratase-Aldolase (tHBPHA). THBPHA is an enzyme which catalyses the final step of the naphthalene degradation pathway. Naphthalene is a fused polycyclic aromatic hydrocarbon that act as a building block for azo dyes and other industrial chemicals. They are often part of industrial contamination into aqueous ecosystem (Cerniglia 1984). There are bacteria (such as Pseudomonas) that have evolved pathways to degrade environmental naphthalene, allowing for naphthalene to become a useable carbon source (Cerniglia 1984).

The naphthalene degradation pathway takes naphthalene of one of its derivatives (such as naphthalenesulfonate) and proceeds to metabolize naphthalene. This results in the reaction of trans-o-hydroxybenzylidenepyruvate which undergoes an aldol cleavage to form pyruvate and salicylaldehyde (Eaton and Chapman 1992).

2.1.1. The final step of naphthalene degradation

The tHBPHA protein catalyses the final step of the naphthalene degradation pathway (Kuhm, Knackmuss et al. 1993). Naphthalene is degraded along the pathway until the substrate trans-hydroxybenzylidenepyruvate (tHBP) is reduced, which then undergoes an aldol cleavage by tHBPHA. This reaction is shown with (figure 6) where tHBP is cleaved into salicylaldehyde and pyruvate. Pyruvate is involved in most reactions from members of the DHDPS/NAL subfamily of proteins, suggesting that the reaction mechanism could be similar. This is shown with the reaction mechanism involving the binding of the highly conserved lysine to the substrate, forming a Schiff-base.

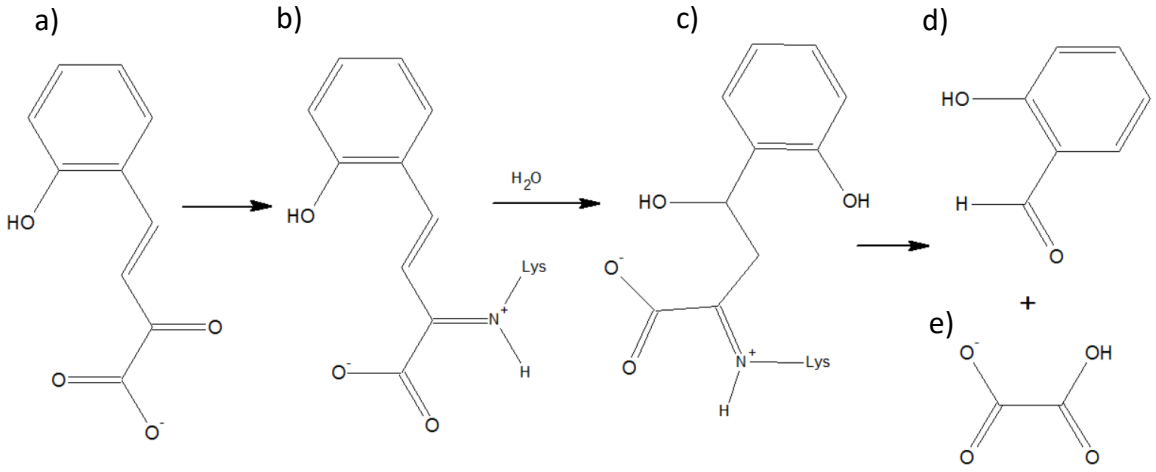


Figure 6

The final reaction of the naphthalene degradation pathway is catalysed by tHBPHA. The substrate tHBP (a) binds within the active, forming a Schiff base with K183 (b), the double bond is then hydrolysed (c) before substrate cleavage into salicylaldehyde (d) and pyruvate (e).

2.1.2. Sequence alignments of members of the subfamily with tHBPHA

In order to investigate tHBPHA with regards to the other members of the DHDPS/NAL subfamily, a multiple sequence alignment was produced for members of the DHDPS/NAL subfamily with multiple tHBPHA proteins (Sievers, Wilm et al. 2011). The proteins included in the alignment are: *Pseudomonas fluorescens* tHBPHA*, Sphingobium xenophagum* tHBPHA*, Escherichia coli* DHDPS and *E. coli* NAL. The crystal structure of tHBPHA from *P. fluorescens* was produced with the work within this thesis and the secondary structure is shown along the top of the alignment. As expected, there are multiple similarities between the members within the DHDPS/NAL subfamily of proteins.

One point is a difference between DHDPS and tHBPHA within the active site. The key active site residues for *E. coli* DHDPS are T44, K161, Y133 and Y107 (Laber, Gomisruth et al. 1992). The sequence alignment shows that the key amino acid residues present within the *E. coli* DHDPS active site are mostly present in tHBPHA from *P. fluorescens*. These are T44 with tHBPHA T65, Y133 with tHBPHA Y155 and K161 with tHBPHA K183. The main difference with the two active sites is that DHDPS Y107 is replaced with W128. K161 performs a nucleophilic attack with its amino group to the keto group of the pyruvate (Blickling, Renner et al. 1997). Y133 of DHDPS is in involved in Schiff base formation with pyruvate and acts to help stabilise the initial reaction. Both enzymes involve use of the substrate pyruvate in their respective catalysis step, which is why these amino acids are present within the active site of both members. Y107, T44 and Y133 are involved in the formation of a catalytic triad, which is important in maintaining the activity of DHDPS, with Y107 also being involved with Y106 in the formation of a hydrophobic stack which is involved in enzymatic inhibition when lysine is present (Blickling, Renner et al. 1997). The lack of Y107 in tHBPHA is due to there being no need to inhibit the enzyme with lysine as in DHDPS but the tryptophan group might still act in a similar way to tyrosine, resulting in possible DHDPS activity.

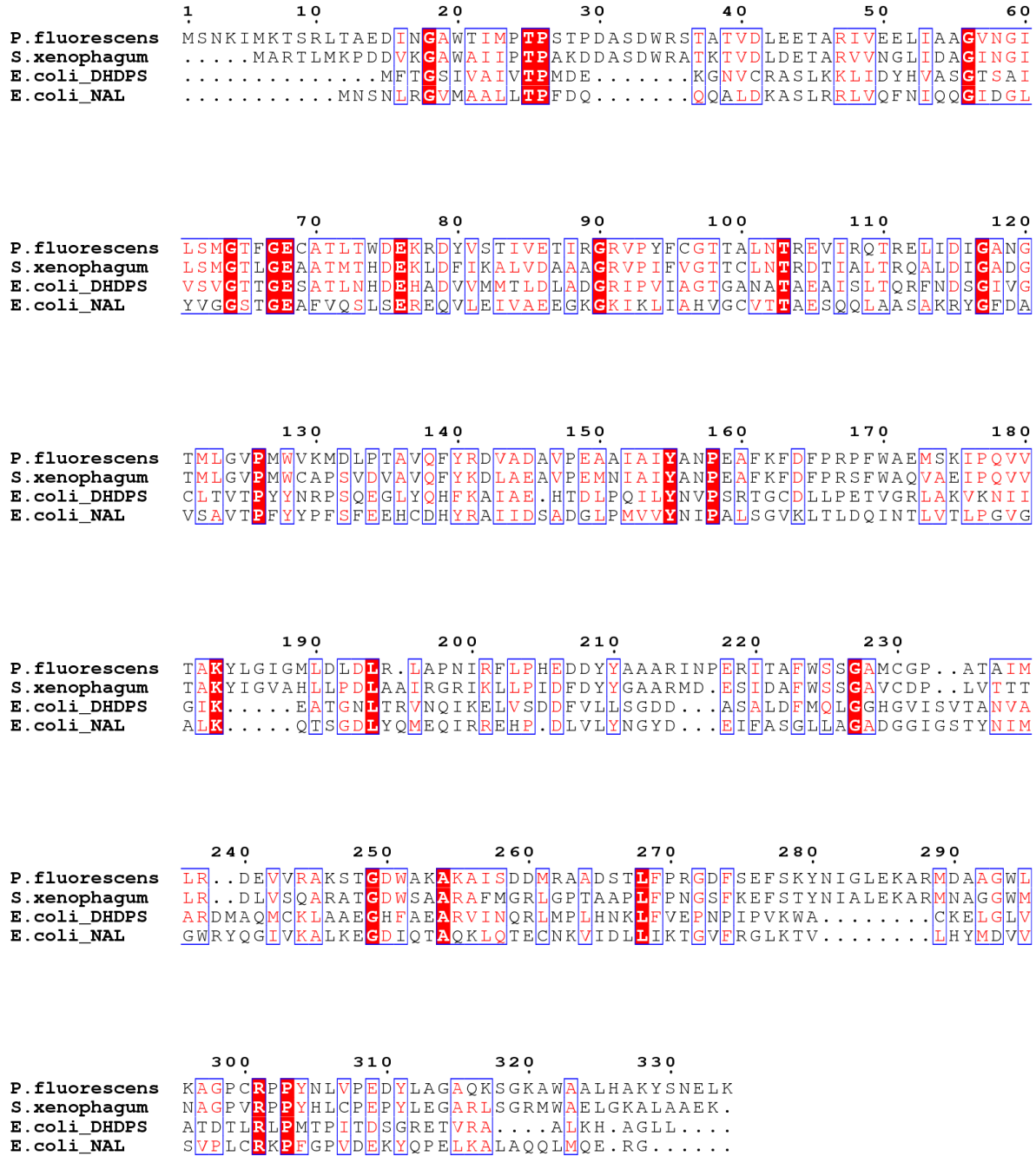


Figure 7

This figure shows the sequence alignments of members of the DHDPS/NAL subfamily of proteins. In descending order*: P. fluorescens* tHBPHA*, S. xenophagum* tBHPHA*, E. coli* DHDPS and *E. coli* NAL. This information is displayed using (escript3.0) software (Robert and Gouet 2014) with secondary structure of *P. fluorescens* tHBPHA being shown along the top. Highly conserved residuals throughout most proteins are shown in red surrounded by a blue box while residuals conserved throughout all members are highlighted by a red background.

2.1.3. Possible other enzymatic activity

Due to the similarities between the structures of members of the DHDPS/NAL subfamily, the possibility for the enzymes to be promiscuous and participate in multiple enzymatic pathways could occur. (Joerger, Mayer et al. 2003) performed specific point mutations on *E. coli* NAL scaffold which resulted in a switch in activity towards DHDPS. While the total sequence identity was below 24% between the two, the overall three dimensional shape is very similar (Joerger, Mayer et al. 2003). They found that one point mutation (L142R) in *E. coli* NAL was sufficient to significantly increase the activity of the mutant for the DHDPS substrate ASA. A 19 fold increase in activity towards this ASA when compared to the residual activity of the NAL wildtype (Joerger, Mayer et al. 2003).

An *E. coli* DHDPS with a mutation in Tyr107 to become phenylalanine to remove the hydroxyl group from the active site resulted in a tenfold loss in activity. This loss (while significant) wasn't as large of loss as other mutants, it was speculated that water molecules were able to pass through the dimer interface towards the phenylalanine mutant, resulting in a water molecule in the position similar to the hydroxyl of tyrosine (Dobson, Valegard et al. 2004, Griffin, Dobson et al. 2008). The tryptophan in the tHBPHA active site could share a similar effect, resulting in DHDPS activity.

The tHBPHA aldol activity was initially assayed by for the first time by (Ferrara et al) who found that the apparent K_m and v_{max}values for the his tagged P. fluorescens tHBPHA was the same (for both its natural substrate from the degradation pathway and for benzaldehyde) (Ferrara, Mapelli et al. 2011). (Sello and Di Gennaro 2013) also assessed the ability of P. fluorescens tHBPHA to use a variation of different aldehydes as a reaction substrate. They found that while being strictly pyruvate depended, the use of different aldehydes as acceptors depending on the electronic characteristics of the aldehyde (Sello and Di Gennaro 2013). The fact that tHBPHA can use different aldehydes as a substrate suggests that a DHDPS activity could be possible. Paired with the high similarity with conserved residues within the active site this becomes even more possible.

2.1.4. Previous structural characterisation of tHBPHA

A previous study characterised tHBPHA from Pseudomonas uesicularis in the 1990's. Using gel filtration, they estimated the molecular weight to be about 120 kDa. Further using gel electrophoresis to estimate the molecular weight of the tHBPHA monomer. They estimated a weight of about 38 kDa, and concluded that he presence of a trimeric protein(Kuhm, Knackmuss et al. 1993). This is an interesting point due to members of the DHDPS/NAL subfamily existing primarily as tetramers. If this oligomeric state is correct then the presence of a novel member of this subfamily has appeared. No apparent structural characterisation of tHBPHA has been taken.

2.1.5. Aims of this chapter

The aim of this chapter is to investigate the structural characteristics of tHBPHA. This will involve investigating the oligomeric state in order to determine if it is a trimeric or tetrameric protein. Techniques such as: Analytical ultracentrifuge (AUC), Small angle xray scattering (SAXS), xray crystallography, will probe the structural properties of tHBPHA while Differential scanning flurometry (DSF) and kinetic assaying will investigate possible DHDPS activity.

2.2. Results

2.2.1. Purification of the tHBPHA proteins

The two tHBP-HA proteins are purified to begin the structural investigation of the family. Two proteins are purified from the species *P. fluorescens (Ferrara, Mapelli et al. 2011)* and *S. xenophagum*. Both proteins were transformed into BL21 (DE3) cells and were grown and purified as described in the methods using His tag purification followed by TEV cleavage and SEC chromatography. *P. fluorescens* tHBPHA showed high levels of expression with ~20-30 mg of protein being expressed from 1.6 L of media. *S. xenophagum* tHBPHA showed a much lower level of expression with ~2-5 mg of protein purified from 1.6 L of media. *E. coli* DHDPS and NAL showed high expression levels with purified protein exceeding 40+ mg/ml from 1.6 L of media.

2.2.2. Analytical Ultracentrifugation of tHBPHA

2.2.2.1. Sedimentation of tHBPHA

The four proteins were analysed and show a similar sedimentation coefficient. The experiment showed *E. coli* DHDPS with a sedimentation value of 7.25 S, compared to the values of *P. fluorescens* tHBPHA of 7.0 S and *S. xenophagum* tHBPHA of 7.1 S. All three proteins show a similar sedimentation suggesting that the oligomeric state is similar. This suggests that the tHBPHA proteins are tetrameric not trimeric. *E. coli* DHDPS and *S. xenophagum* tHBPHA seem to possibly have another species present at 4.5-5 S which could be in line with a dimer or trimer. Its presence shouldn't be large enough to influence the analysis due to the significant difference between the two peak heights.

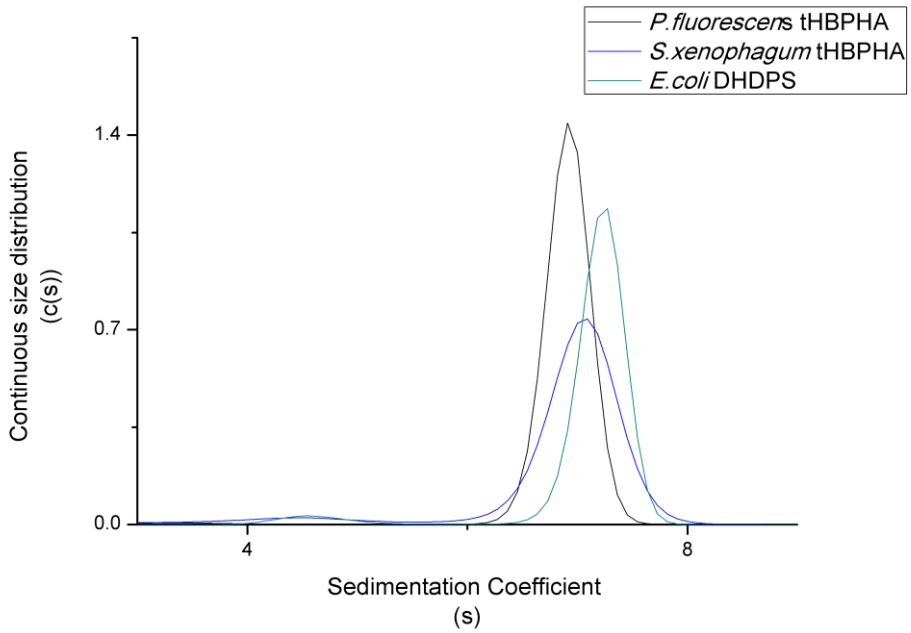


Figure 8

AUC Sedimentation velocity experiments with; 0.5 mg/ml *E. coli* DHDHPS, 0.4 mg/*ml P. fluorescens* tHBPHA and 0.2 mg/ml *S. xenophagum* tHBPHA. DHPS is used as the reference for a tetrameric protein with both tHBPHA proteins exhibiting similar sedimentations.

The tetrameric state isn't influenced by the concentration of tHBPHA. To determine if the oligomeric state of tHBPHA is influenced by protein concentration, AUC sedimentation velocity was used with varying concentrations of *P. fluorescens* tHBPHA (figure 9). The experiment used varying concentrations of tHBPHA at 0.1 mg/ml, 0.2 mg/ml and 0.4 mg/ml. All three shared a similar sedimentation coefficient of ~7 S with no other species being present suggesting only one oligomeric state is present.

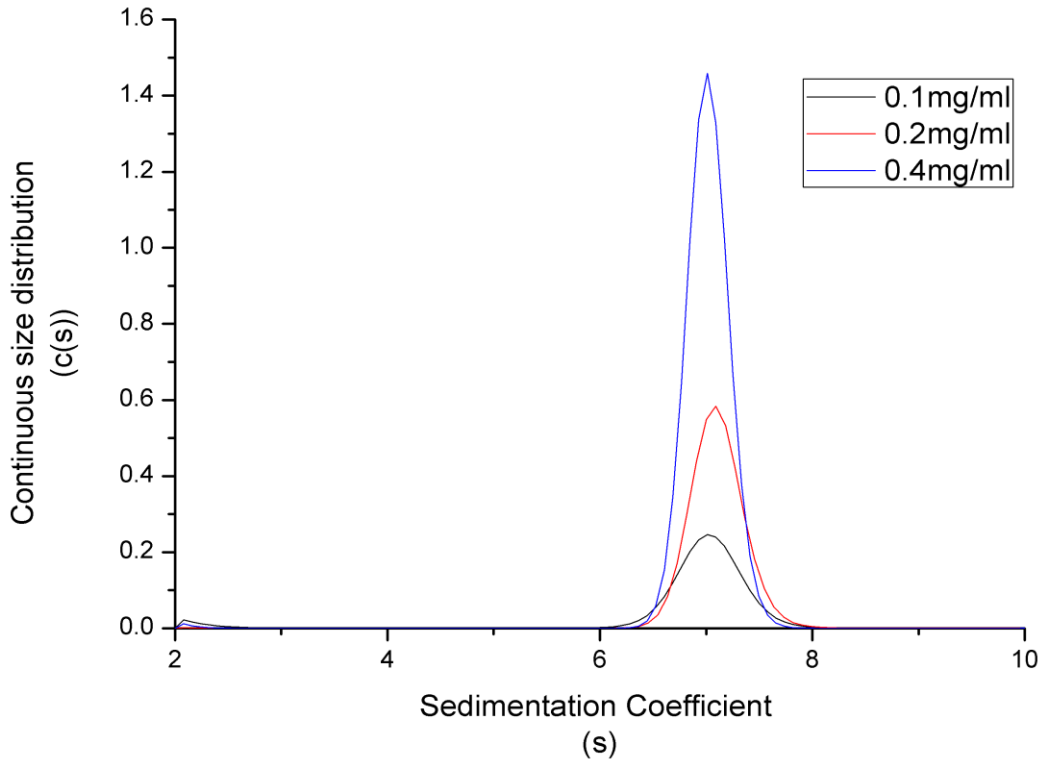


Figure 9

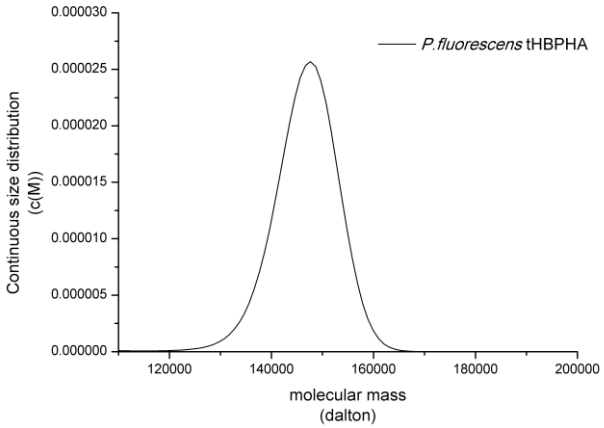
AUC sedimentation velocity of three different concentrations of *P. fluorescens* tHBPHA; 0.1 mg/ml, 0.2 mg/ml and 0.4 mg/ml. This shows that the tetrameric orientation isn't concentration depended as at no point does the presence of a dimer / trimer appear.

2.2.2.2. Molecular Mass of the tHBPHA Proteins

Figure 10 shows the molecular weights both *P. fluorescens* and *S. xenophagum* tHBPHA as calculated through AUC sedimentation velocity with both showing a molecular mass corresponding with a tetrameric protein. The theoretical Molecular weight of *P. fluorescens* tHBPHA monomer is 36.9 kDa, giving a theoretical trimeric weight of 110.826kDa and a tetrameric weight of about 147.8 kDa. The molecular mass for *P. fluorescens* tHBPHA determined using AUC is 147.6 kDa (figure 10, a)). There is only a slight difference between the theoretical and the calculated molecular masses, showing that this protein is in a tetrameric orientation.

S. xenophagum tHBPHA shows a theoretical monomer weight of 35.5 kDa with a trimeric weight of 106.4 kDa and a tetrameric weight of 141.8 kDa. The molecular mass for *S. xenophagum* tHBPHA was calculated at 145.6 kDa (figure 10, b)). As with *P. fluorescens* tHBPHA, this value is consistent with a tetrameric protein when compared to the theoretical tetrameric weight.

a)



b)

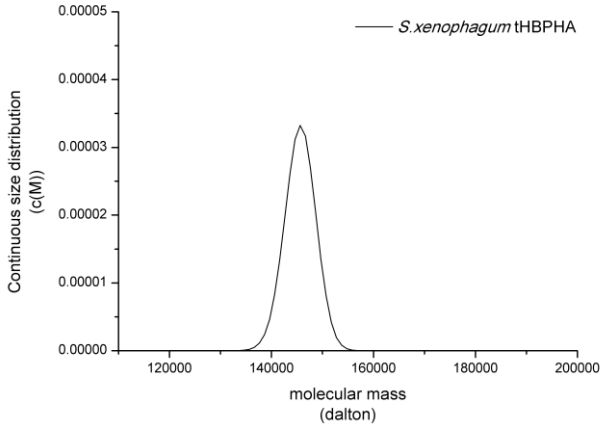


Figure 10

AUC sedimentation velocity data showing the calculated molecular mass of the tHBPHA proteins. These graphs show P. fluorescens tHBPHA (a) and S. xenophagum tHBPHA (b). Both proteins are consistent with a tetrameric mass.

2.2.3. Small angle x-ray scattering

2.2.3.1. Intensity Data

In order to determine the solution structure of tHBPHA proteins, small angle x-ray scattering (SAXS) is employed. Samples of *P. fluorescens* and *S. xenophagum* tHBPHA were examined using SAXS along with tetrameric *E. coli* DHDPS and NAL as references (figure 11).

E. coli DHDPS and NAL were used as a reference for a tetrameric proteins as well as a quality control to ensure that the weights aren't influenced by on site issues. The calculated weight from SAXS showed 139.3 kDa for *E. coli* DHDPS which has a theoretical tetrameric weight of 125 kDa. SAXS data for *E. coli* NAL estimated the molecular weight of 134.527kDa which is comparable to the theoretical tetrameric weight of 130.5 kDa.

P. fluorescens tHBPHA produced a porod volume of 203406. The zero scattering angle (I_0) value is 0.071, the radius of gyration (R_g) is 33.48. The calculated molecular weight produced from SAXS is about 141 kDa. When compared to the theoretical sequence weights, the result is closer to the theoretical tetrameric weight of 147.8 kDa, showing the presence of a tetrameric protein.

S. xenophagum tHBPHA showed a porod volume of 178111, an I_0 of 0.04 and an R_g of 35.01. SAXS produced a molecular weight of about 126.4 kDA protein. This is somewhat comparable to both the theoretical tetramer (141.8 kDa) and theoretical trimer (106.3 kDa). This result also suggests that a tetrameric protein is likely to be present. The large difference between both tHBPHA proteins could be due to quality of *S. xenophagum* tHBPHA or protein concentration.

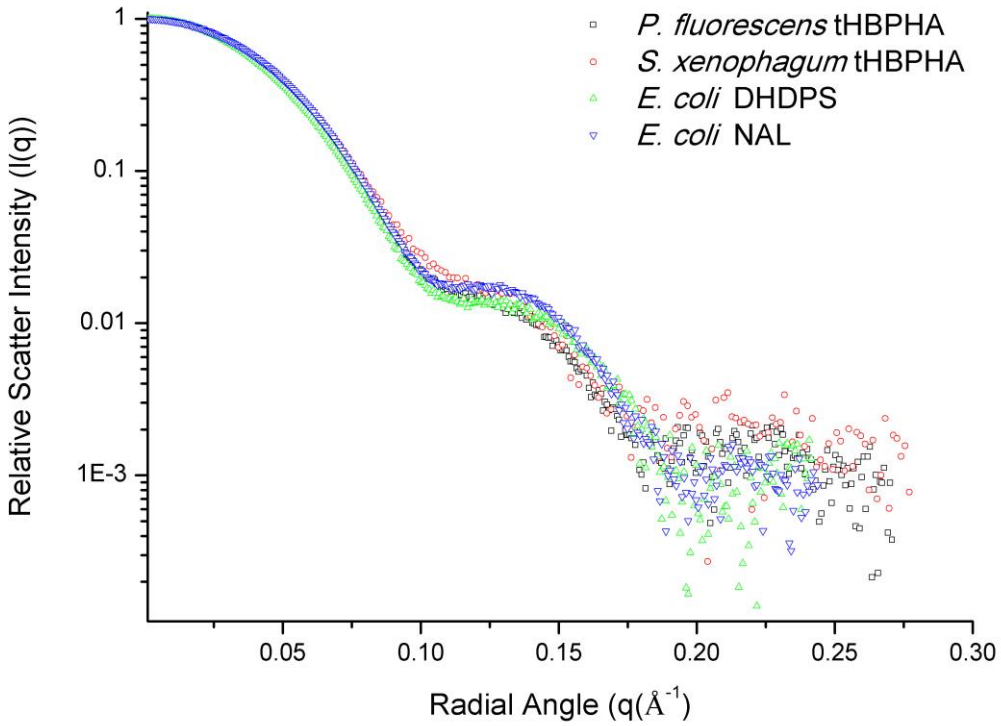


Figure 11

SAXS data for intensity plots are shown with *E. coli* DHDPS, *E. coli* NAL, *P. fluorescens* tHBPHA and *S. xenophagum* tHBPHA. The four plots are normalised and overlaid showing the similarities in SAXS scatter due to the similarities between the states of the solution structures.

2.2.3.2. p(r) Distribution Plots

(Figure 12) shows a distance distribution plot of the four proteins. From the calculated molecular weights we can see that all four proteins exhibit a similar oligomeric state. Four protein p(r) plots are normalised and aligned this shows that both tHBPHA and the tetrameric reference proteins of *E. coli* DHDPS and NAL all share a similar p(r) distribution resulting from a similar presence in solution. This shows that, within solution, all four proteins exist in a similar state. This state is similar to the state that DHDPS/NAL exists as or the p(r) plot would show a different distribution. With the exception of *E. coli* DHDPS, the maximum partial size of both P. fluorescens and *S. xenophagum* tHBPHA as well as *E. coli* NAL is similar resulting from a similar particle size in solution.

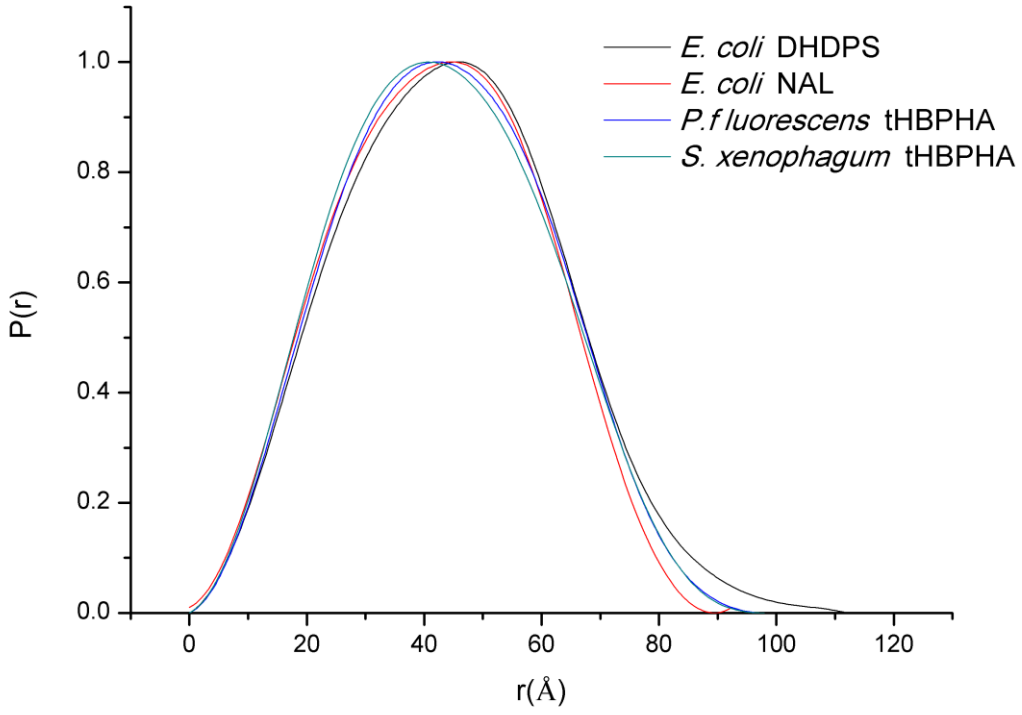


Figure 12

The p(r) distributions for *E. coli* DHDPS, *E. coli* NAL, *P. fluorescens* tHBPHA and *S. xenophagum* tHBPHA. All four p(r) distributions are overlaid showing a similar distribution resulting from existing in a similar state in solution. The maximum dimensions of the proteins are also consistent (with exception of DHDPS).

2.2.3.3. Kratky plots showing folded protein

Kratky plots are calculated from intensity data and are used to determine data quality and how well folded the suspect protein is (Putnam, Hammel et al. 2007). Figure 13 shows all four Kratky plots for: *P. fluorescens* tHBPHA, *S. xenophagum* tHBPHA, *E. coli* DHDPS and *E. coli* NAL. All four proteins show a similar Kratky plot with a single, uniform peak between 0-0.1 Å and scatter past 0.25 Å. This shows that all present proteins are folded within solution.

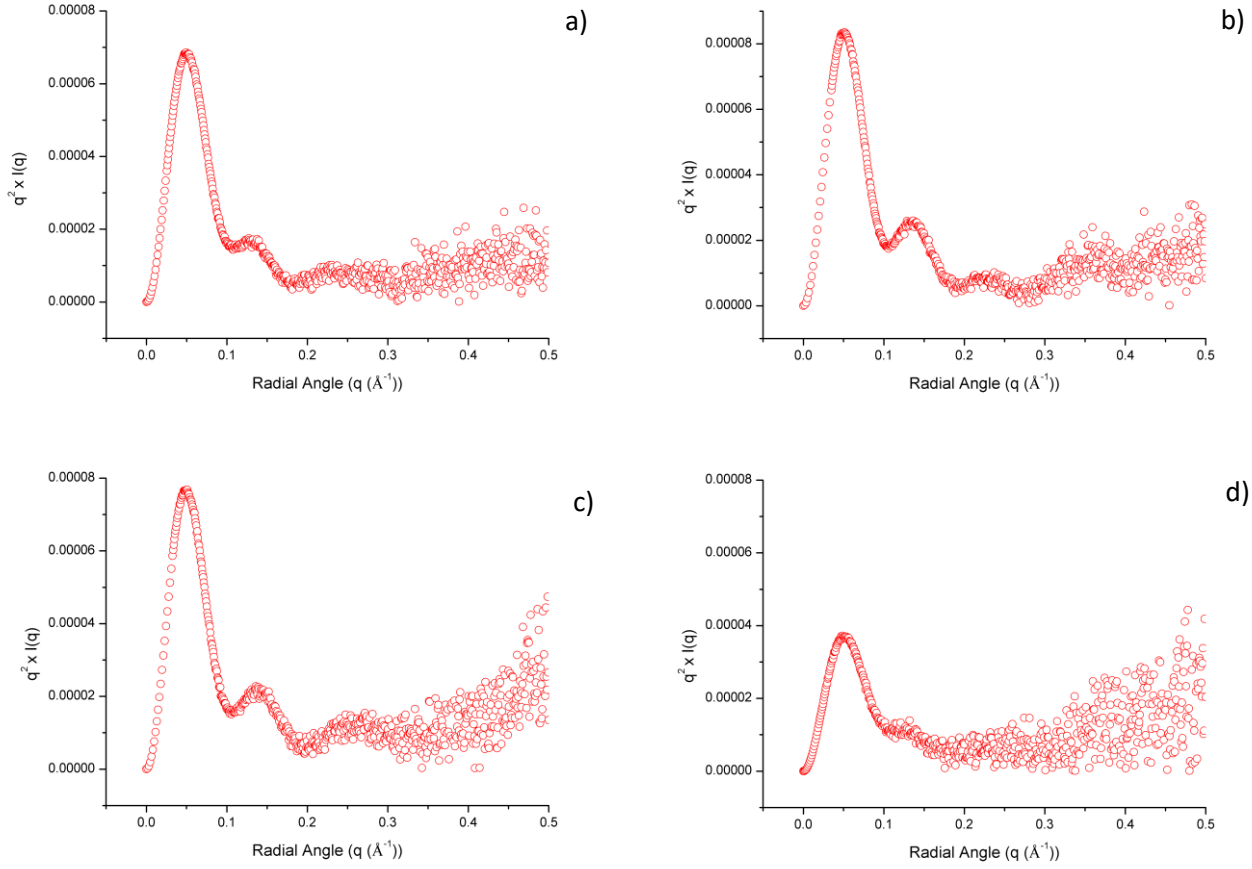


Figure 13

SAS data was modelled in a Kratky plot, the figures show: *E. coli* DHDPS (a), *E. coli* NAL (b), *P. fluorescens* tHBPHA (c) and *S. xenophagum* tHBPHA (d). The smooth peak frim 0-0.1Å shows that each protein is globular within solution as opposed to fold.

2.2.4. X-ray crystallography of *P. fluorescens* tHBPHA

2.2.4.1. Solving the crystal structure

In order to further probe the active site architecture to investigate the structural consequences of the alternate residues, x-ray crystallography trials were conducted. From the PACT suite, the condition which produced the best crystal diffraction data was well G7 (0.2M Sodium malonate dibasic monohydrate with 20% w/v PEG 3350) which diffracted to 2.2 Å. Molecular replacement was done using AUTO-RICKSHAW from the Australian synchrotron (Panjikar, Parthasarathy et al. 2005, Panjikar, Parthasarathy et al. 2009). Initially, molecular replacement was unsuccessful but through production of a second, lower resolution crystal in the space group I4122.

Auto Rickshaw initially used Pirate, a programme which statistically modifies the phase probability distributions from experimental phasing to produce an improved version of the phase probability distributions to aim to calculate a clearer map (Cowtan 2000). This followed with the use of SHELXE for further phase and density modification (Sheldrick 2010). Auto Rickshaw then used the program buccaneer to produce a model suitable for molecular replacement. Buccaneer is automated program used for protein model building (Cowtan 2006, Cowtan 2008). This was done effectively by 'growing' a backbone within the electron density following use of probability functions to predict the orientation of the side chains and where they sit with the density. (Cowtan 2006, Cowtan 2008)Through use of AUTO-RICKSHAW (Panjikar, Parthasarathy et al. 2005, Panjikar, Parthasarathy et al. 2009) and Buccaneer (Cowtan 2006, Cowtan 2008), a molecular replacement model was produced and structure refined. This was then used to solve the original diffraction pattern. The r-word was refined down until 0.185 while the r-free was refined to 0.235.

2.2.4.2. tHBPHA crystal structure is in a tetrameric orientation

2.2.4.2.1. Monomeric structure of tHBPHA.

The structure of tHBPHA was solved to 2.2 Å. Its structure exhibits similarities to that of DHDPS/NAL with the highly conserved TIM barrel fold being represented in the tHBPHA family. This is shown with (Figure 14, a)), the monomer consists of eight β-sheets surrounded by eight α-helices as is the standard for members of the DHDPS/NAL subfamily of proteins. There are a further three α-helices which aren't a part of the barrel by are present to the side. (Figure 14, b)) shows the inverted monomer with the presence of a small loop which resides below the barrel.

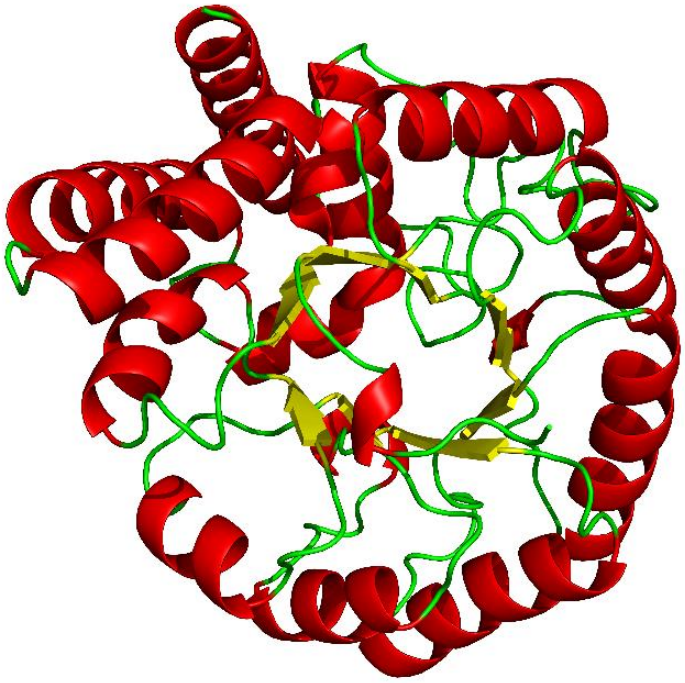


Figure 14

The crystal structure of the *P. fluorescens* tHBPHA monomer is reported with: α-helices are shown in red, are β-sheets shown in yellow and the remaining chain in green. The structure consists of eight β-sheets and twelve α-helices. The active site is present within the barrel similar to other TIM barrel proteins.

2.2.4.2.2. Tetrameric Structure of tHBPHA

The tHBPHA structure is shown to exist as a tetrameric protein (as supported with AUC and SAXS results) which is consistent with that of DHDPS/NAL. Similar to DHDPS and NAL, tHBPHA seems to assemble as a dimer of dimers. This is expected due to tHBPHA being a member of the same subfamily. The homo-tetrameric structure of tHBPHA is shown in (Figure 15), the active site is within each barrel and involves interactions with the opposing dimer (assuming a similar reaction mechanism to DHDPS).

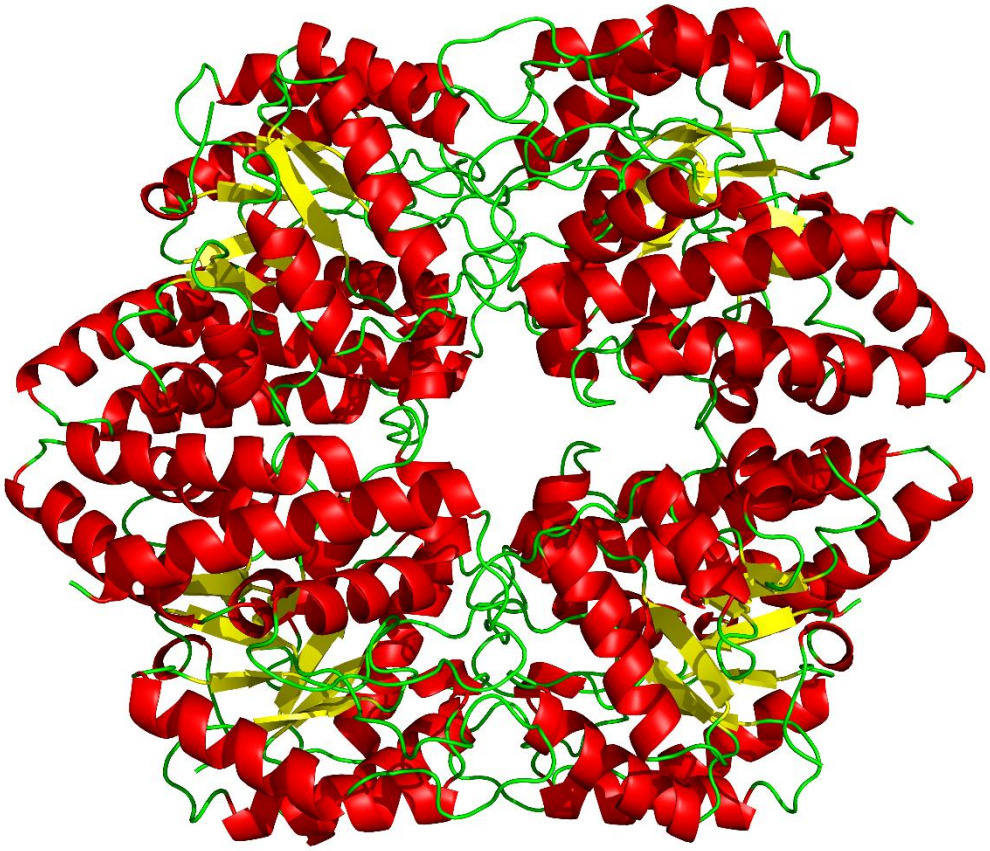


Figure 15

The tetrameric orientation of tHBPHA from *P. fluorescens* is shown with: α-helices are shown in red, are β-sheets shown in yellow and the remaining chain in green. The structure fits the tetrameric orientation well and doesn't seeming show the possibility for a trimeric orientation.

2.2.4.2.3. The tHBPHA theoretical crystal scatter compared to solution scatter produced from SAXS

To probe if the tetrameric crystal structure is consistent with that of the solution structure determined with SAXS, the theoretical scatter from the tHBPHA crystal structure was compared to the scatter produced from the solution with SAXS. The program Crysol (Svergun, Barberato et al. 1995) was used to compare the scatter as shown with Figure 16. The figure shows the theoretical scatter of the tetrameric structure (figure 16, a)) and dimeric structure (figure 16,b)) which is compared to the actual tHBPHA scatter from SAXS. A low Chi^2 value of 0.215 was produced for the tetrameric structure while a much higher Chi^2 value of 62.667 was produced from the dimeric structure. The significant difference in the chi^2 values between the two structures show that the tetrameric structure is present within the solution. This further confirms that the tetrameric structure for tHBPHA is the correct orientation.

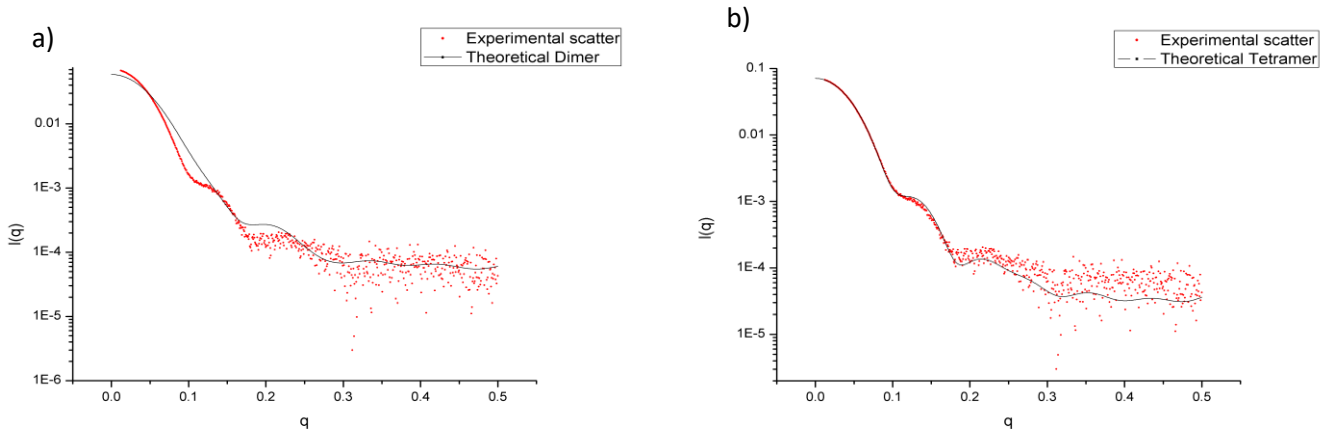


Figure 16

This figure shows the theoretical scatter from the *P. fluorescens* crystal structure (black line) and compares it to the scatter (red dots) from the solution structure to determine if the crystal structures oligomeric state is valid in solution. Figure a) shows a dimeric tHBPHA while the tetrameric orientation is shown in b). The dimeric structure produced a chi^2 value of 63.33 while the tetrameric structure produced a chi^2 value of 0.215.

2.2.4.2.4. Speculation of the tHBPHA Active site.

The active site of tHBPHA is highly conserved from the DHDPS/NAL subfamily of proteins. They share a similar active site with regards to what residues are present within the active site. Whilst the crystal structure hasn't been produced with substrate bound in the active site, due to the similarities between members of the DHDPS/ NAL subfamily, the amino acids that are involved in catalysis can be speculated.

Due to the sequence similarities between members of the DHDPS/NAL subfamily, the key amino acids which are conserved are likely to be present across the members. This is true in the active site of tHBPHA and is clearly seen when looking at the sequence alignment (figure 7). The three residues that are conserved are; T65, Y155, K183. The different between DHDPS and tHBPHA is the presence of W128 instead of tyrosine in DHDPS. Y107 in *E. coli* DHDPS is important for catalytic function due to its involvement as part of the catalytic triad as well as hydrophobic stacking with the opposite tyrosine (Blickling, Beisel et al. 1997, Dobson, Griffin et al. 2005). Y107 in DHDPS also acts to form a hydrophobic stack with Tyrosine 106 within the active site (Reboul, Porebski et al. 2012). While tHBPHA can use W128 to mimic the hydrophobic tyrosine in DHDPS, there isn't another aromatic amino acid group next to the tryptophan in which to stack with. The presence of F161 however is in close proximity to the tryptophan which could allow for a hydrophobic stacking effect to occur.

As with DHDPS, the active site will involve the participation of the opposite monomer of tHBPHA. W128 can extend into the opposing active site in a similar manner to how Tyrosine does in DHDPS. This is shown with (Figure 17) where the first chain in green (showing T44, Y155, and K183 within the active site) also contains W128 which extended form the second chain in red.

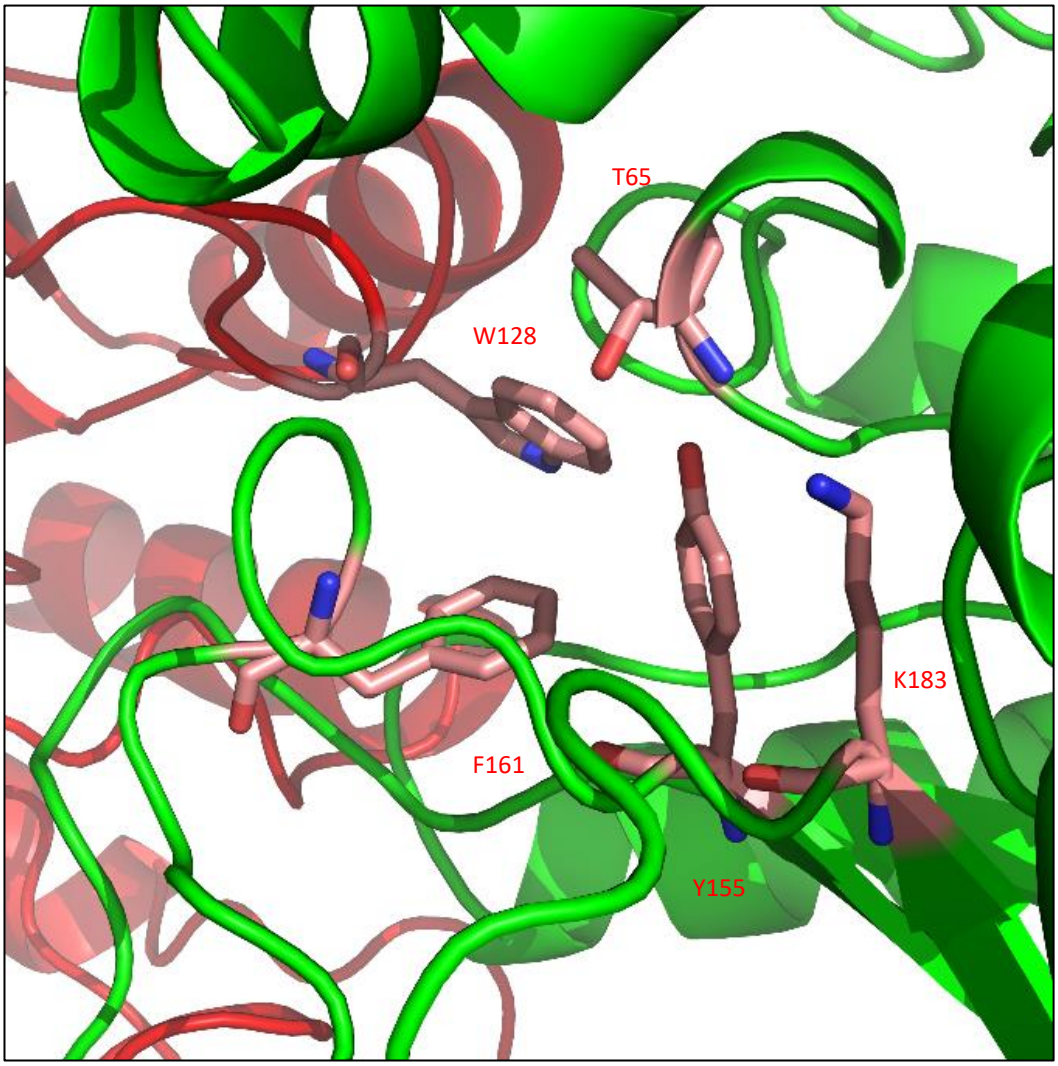


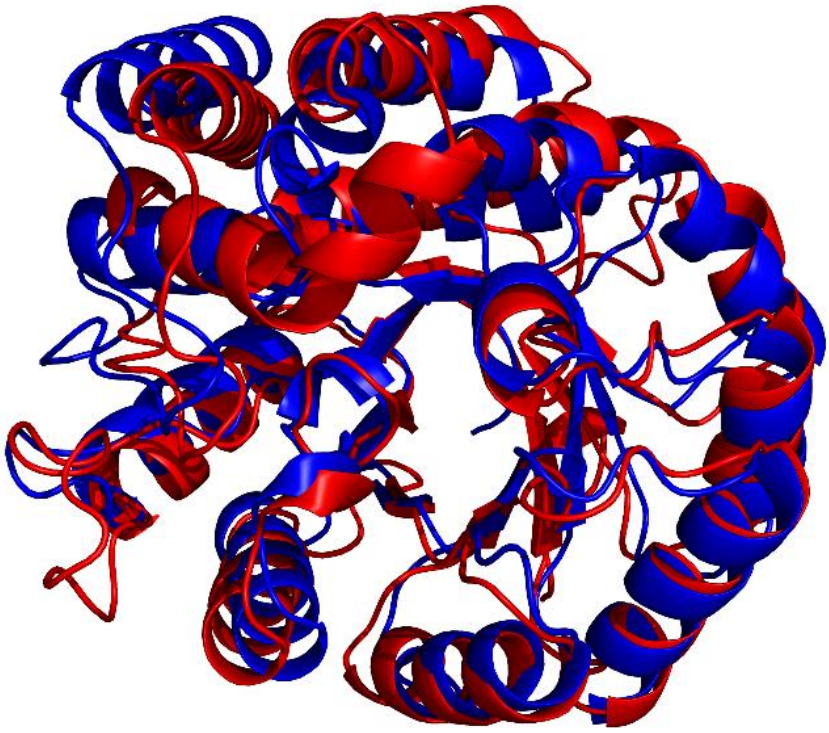
Figure 17

The key residues of *P. fluorescens* tHBPHA are shown. Lysine is involved in binding of pyruvate and can face towards the other amino acids within the active site. Threonine and tyrosine could interact with tryptophan within the active site similar to how tyrosine does in DHDPS. Tryptophan inserts itself into the active site from the opposing monomer. Phenylalanine could act with W128 to form a hydrophobic stack.

2.2.4.3. Alignment of tHBPHA to *E. coli* DHDPS

2.2.4.3.1. Alignment of two Monomers

The monomer of tHBPHA was overlaid with the monomer of *E. coli* DHDPS (1YXC) (Dobson, Griffin et al. 2005) as shown in (Figure 18) The TIM barrel fold is extremely similar with only some slight variation between them which is due to both proteins belonging to the same subfamily. There is variation in the other α-helices not present in the TIM barrel. The tHBPHA structure also shows more random protein loops than the DHDPS structure.



POA_DHDPS_ALIGNMENT

The monomeric structure of tHBPHA was aligned to that of *E. coli* DHDPS (1YXC) (Dobson, Griffin et al. 2005) from a view looking down on the monomer. There is a slight variation between the two proteins with regards to the barrel but it exhibits a high level of similarity as expected form a member of the same subfamily. The main variation between the two proteins is due to the additional helices to the side of the barrel.

2.2.4.3.2. Alignment of the tHBPHA and DHDPS active sites

Figure 19 shows the active sites of tHBPHA and *E. coli* DHDPS aligned. They both share similarities in where the amino acids are positioned. The lysine of tHBPHA seems to be in a different position from the DHDPS counterpart. This is likely due to the lack of a substrate present to interact with the lysine group. If pyruvate were to be present in the tHBPHA structure, K183 would likely be facing toward the other amino acids in the active site. T65 and Y155 are in a very similar orientation and position to their DHDPS counterparts. Tryptophan is positioned in a similar position to the DHDPS tyrosine but its catalytic role in the active site is unknown or speculation.

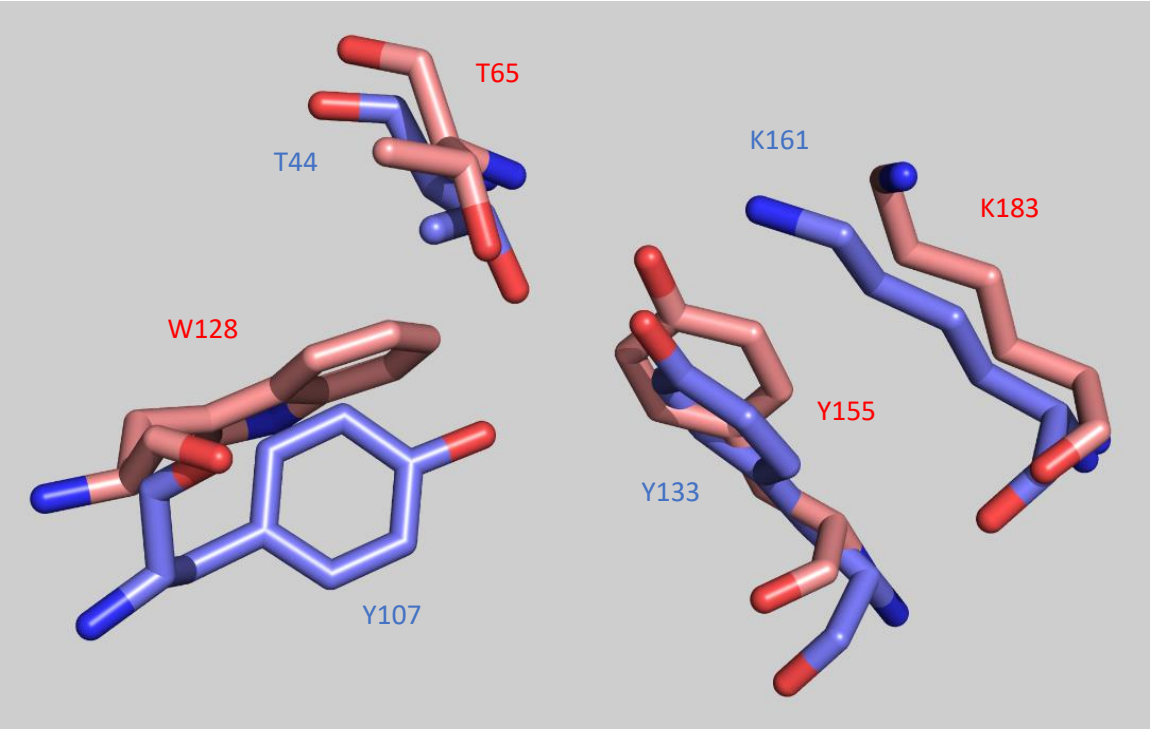


Figure 19

The active site residues for *E. coli* DHDPS (blue) and *P. fluorescens* tHBPHA (red) are overlaid. Tyrosine, lysine and threonine of both DHDPS and tHBPHA are in similar orientations. The lysine of tHBPHA would point towards the other amino acids in presence of pyruvate. Y107 in DHDPS is replaced by W128 in tHBPHA.

2.2.5. Differential scanning fluorimetry to investigate substrate binding

Differential scanning fluorimetry (DSF) was used on *P. fluorescens* tHBPHA proteins as well as *E. coli* DHDPS as a reference for a protein which will bind the substrates. Four conditions were used for each protein; Protein, protein in presence of Pyruvate, protein in presence of lysine and protein in presence of both lysine and pyruvate. This experiment is used to determine the relative melting points of each protein as well all to determine if the presence of each substrate could affect the melting points of each protein.

E. coli DHDPS (a) showed an initial T_m of 60.82^{o}C, this increased to 62.74^{o}C with incubation in pyruvate, 65.12^{o}C with lysine and 66.44^{o}C with incubation of both substrates. For *P.fluorescens* tHBPHA (b), the relative melt points (T_m) for the protein was 51.23^{o}C, this increased to 51.88^{o}C with addition of pyruvate, 50.86^{o}C with incubation with lysine and 52.14^{o}C when incubated with both. The result shows a slight increase in the melting temperature of the protein which suggests that binding of substrate could be occurring. If this change in T_m is significant requires further investigation.

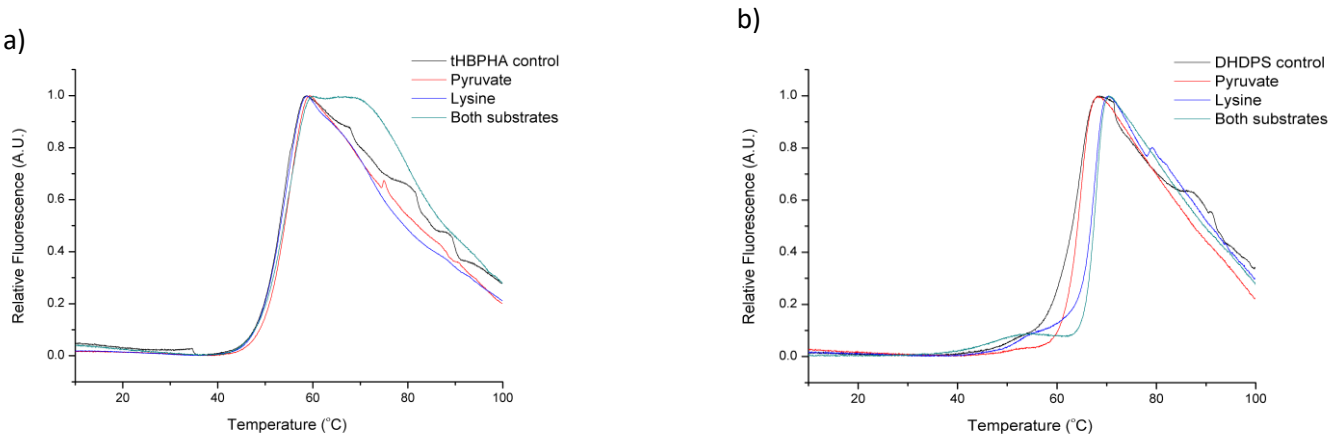


Figure 20

Thermal shift assay to examine the ability for tHBPHA to bind substrates related to the DHDPS. *P. fluorescens* tHBPHA (a) and *E. coli* DHDPS (b). There is no significant change with the addition of substrates, suggesting that tHBPHA isn't binding pyruvate or lysine as well as DHDPS.

2.2.6. Kinetic assay for DHDPS activity

To investigate if tHBPHA capable of acting in place of DHDPS to catalyse the first step of the DAP pathway, a kinetic assay for DHDPS activity was investigated. Due to the sequence similarities between tHBPHA and DHDPS as well as the similarities within the active site, a DHDPS kinetic assay was formed using tHBPHA in place of DHDPS. There was no significant activity of tHBPHA when added into the DHDPS assay. This suggests that tHBPHA is unable to catalyse the DHDPS.

2.3. Discussion and Conclusion

The purifications of the two different tHBPHA proteins had mixed results due to the variation between protein yields from the two proteins. tHBPHA from *P. fluorescens* produced high yields of proteins during purifications (~30 mg or more of protein per purification from 1.6L of media) while tHBPHA from *S. xenophagumi* showed far lower yields (2-5 mg per purification from 1.6L of media). This resulted in difficulties in examining *S. xenophagumi* tHBPHA due to the large differences in protein present while *P. fluorescens* tHBPHA was able to be well characterised.

Another tHBPHA was to be examined within this thesis but the yield and purity produced was too low to be accurately produce useful data from analytical techniques. The quality of *S. xenophagumi* tHBPHA was also lower than that of *P. fluorescens,* resulting in a higher level of difficulty to obtain similar results.

One challenge in obtaining the crystal structure of tHBPHA is due to the lack of a compatible molecular replacement model due to a low sequence identity between other solved structures and tHBPHA. In spite of DHDPS and NAL having numerous solved structures, there are currently a lack of solved structures for tHBPHA. The closest sequence identity to *P. fluorescens* tHBPHA was: DHDPS from *Acinetobacter baumannii* at 31.36% identity,

DHDPS from Pseudomonas aeruginosa 27.98% identity and DHDPS from *Legionella pneumophila* 28.47% identity (Altschul, Madden et al. 1997, Altschul, Wootton et al. 2005). These proteins were analysed using PHENIX (Afonine, Grosse-Kunstleve et al. 2012) to attempt molecular replacement but were unsuccessful because of this low identity.

With the production of a suitable model through Buccaneer, the *P. fluorescens* tHBPHA crystal was able to be solved through use of molecular replacement. If this wasn't the case then the structure would, likely, have to be solved through production of Selenomethionine crystals in order to solve the structure through use of multi-wavelength anomalous diffraction (MAD) measurements (Wu, Lustbader et al. 1994). With a suitable model for molecular replacement being produced, the crystal structure could then be solved. Refinement was able to bring the initial R-free from 0.38 to 0.235.

2.3.1. tHBPHA is in a tetrameric protein

As was previously described in (Kuhm, Knackmuss et al. 1993), tHBPHA showed the possibility to exist as a trimer within solution. This would represent a novel member of the DHDPS/NAL subfamily of proteins which mostly exist with a tetrameric or dimeric orientation. To investigate the possibility of these proteins existing as a timer, AUC sedimentation velocity was used to determine the oligomeric state and the molecular weights of the tHBPHA proteins.

The AUC data shows tHBPHA from both *P. fluorescens* and *S. xenophagum* sedimentating in a similar way to that of *E .coli* DHDPS as shown in figure 8. The similar sedimentation suggests that all three proteins share a similar weight and shape. This is shown with tHBPHA from *P. fluorescens* showing a sedimentation value of 7.0 S and *S. xenophagum* showing a sedimentation value of 7.1 S. Both values are similar to the sedimentation value of *E. coli* DHDPS at 7.25 S. The similarities between these proteins suggested that they sediment in a

similar way which is likely the result of sharing a similar size and orientation. When the molecular mass was calculated, it further showed that tHBPHA was consistent with a tetrameric mass of about 147.6 kDa for *P. fluorescens* tHBPHA and 145.6 kDa for *S. xenophagum* tHBPHA. These weights are consistent with a tetramer and far out weight the mass of what the theoretical trimer is.

There was the slight presence of a secondary species which could have shown a dimeric/trimeric mass. The possibility that the oligomeric state of tHBPHA could be based on the concentration of protein was also investigated with (figure 9) which aimed to confirm if a change in oligomeric state was due to different protein concentrations. All three different concentrations show the same sedimentation value with no change anywhere else being noted. This shows that the tetrameric orientation of tHBPHA is constant and not based on the concentration of protein present.

To further prove the presence of a tetrameric protein, SAXS was performed to investigate the solution structure. It allows for the validation of AUC results as well as providing a way to also confirm the crystal structure is similar to the structure within solution. Both tHBPHA proteins produced a well organised Kratky plot showing the presence of folded, globular proteins in solution. The intensity plots were also very similar for both the tetrameric *E. coli* DHDPS/NAL and tHBPHA proteins. This suggests that the tetrameric references and tHBPHA are similar. This is further shown with similarities in the p(r) plots. SAXS estimated that the molecular weight of *P. fluorescens* tHBPHA is 141.2 kDa, this value is similar to the one produced from AUC (147.6 kDa) further proving the presence of a tetrameric tHBPHA.

This result is less consistent for *S. xenophagum* tHBPHA when compared to the molecular weights produce from AUC. AUC estimated the molecular mass to be about 145.6 kDa while SAXS calculated the weight to be 126.4 kDa. This value is slightly closer to the theoretical tetrameric structure than the trimeric structure. When coupled with the calculated mass from AUC, it suggests that *S. xenophagum* tHBPHA should be in a tetrameric orientation.

This result is similar to the one obtained from (Kuhm, Knackmuss et al. 1993) who predicted the presence of a trimeric protein. Their prediction of a weight close to 120 kDa is consistent to the SAXS result but is contrasted by the AUC result. While the *S. xenophagum* tHBPHA molecular weight could be interpreted as either trimeric or tetrameric, the *P. fluorescens* tHBPHA is a confirmed tetramer.

The *P. fluorescens* tHBPHA was shown in a tetrameric orientation similar to DHDPS and NAL. Through prediction of the theoretical scatter from the crystal structure and comparison to the SAXS scatter, the authenticity of the crystal structures state can be determined. The chi^2 value from the tetrameric graph is 0.215 which is significantly lower than the dimeric value of 62.667. These values show a clear difference which results in the tetrameric orientation being present in solution over a dimeric one, resulting in the tetrameric crystal structure showing the correct orientation. The crystal structure was also unable to be solved in a trimeric orientation, further showing that the tetrameric orientation is correct.

2.3.2. tHBPHA is similar to members of the DHDPS/NAL subfamily

The previous work has shown that tHBPHA from both *P. fluorescens* and *S. xenophagum* exist in a tetrameric orientation. This has shown that the previous work by (Kuhm, Knackmuss et al. 1993) who reported a trimeric tHBPHA was inaccurate due the gel results. While a trimeric orientation would be an interesting structure for a member of the DHDPS/NAL subfamily due to their need to be present in a dimeric orientation for enzymatic activity, it has been shown that a tetramer is present. The oligomeric state and the presence of the TIM barrel fold resulted in an overall structure which is conserved through this subfamily.

With the solved crystal structure of *P.fluorescens* tHBPHA, the similarities between tHBPHA and DHDPS / NAL can be clearly seen. Members of the DHDPS sub family all share the

characteristic TIM barrel fold. Figure 19 showed *P. fluorescens* tHBPHA aligned with the *E.* coli DHDPS. The conserved amino acids were in similar orientations to DHDPS, which suggested the mechanism of catalysis would follow a similar route. One interesting point is the difference between *E. coli* DHDPS and tHBPHA from *P. fluorescens*. The only major difference between DHDPS and tHBPHA being the tyrosine from the opposite monomer in DHDPS being replaced with a tryptophan in tHBPHA. How tHBP would bind within the active site with regards to the presence of the tryptophan within the active site is a possible point in which future investigation could take place.

2.3.3. DHDPS activity from tHBPHA?

Previous characterisation of tHBPHA from *P. uesicularis* included some kinetic analysis of tHBPHA(Kuhm, Knackmuss et al. 1993). Another found that tHBPHA was able to recognise a number of different substrates which varied on levels of similarity to the original substrate (Sello and Di Gennaro 2013). This shows the possibility for tHBPHA to promiscuously act in the DHDPS reaction due to its ability to catalyse different substrates and therefor showing a possibility to catalyse the DHDPS reaction. This thesis investigated the possibility of tHBPHA exhibiting DHDPS activity through use of a DHDPS kinetic assay as well as DSF to determine possible substrate binding.

First, the use of DSF to determine if pyruvate and lysine were binding to tHBPHA. There was no significant change in the melting point of tHBPHA with the addition of pyruvate or lysine. There was a slight increase with each addition but the increase doesn't seem significant when compared to the large increase from DHDPS.

A DHDPS assay was performed to test the DHDPS activity of tHBPHA from *P. fluorescens*. It was determined that there was no significant activity from tHBPHA. This shows that tHBPHA is unable to perform a role as a promiscuous enzyme and participate in both pathways. It

was speculated that the presence of tryptophan could have performed the role of tyrosine similar to how a phenylalanine mutant did for DHDPS, but that doesn't seem to be the case.

2.3.4. Conclusion

tHBPHA is a tetrameric protein as determined through multiple techniques. AUC determined that both tHBPHA proteins from *P. fluorescens* and *S. xenophagum* are in a tetrameric orientation due to the sedimentation coefficients being similar to the tetrameric DHDPS as well as the molecular weights determine from AUC corresponding with a tetrameric protein. SAXS performed on the tHBPHA proteins suggest that the proteins are in a tetrameric orientation. The tetrameric orientation is proven through the solution structures molecular weights which are calculated through the porod volumes produced from each protein through SAXS. *S. xenophagum* tHBPHA showed an odd molecular weight value from SAXS which was in-between tetramer and trimer.

The crystal structure produced for tHBPHA from *P. fluorescens* also shows the tetrameric orientation is present. This is shown with a comparison of the theoretical x-ray scatter produced from a dimeric and a tetrameric tHBPHA crystal structure and their comparison to the SAXS solution scatter. The tetrameric structure showed a much higher correlation to the solution scatter than the dimeric structure and concludes, again, the presence of a tetrameric structure. The active site is similar to DHDPS with T65, Y155 and K183 being present in DHDPS by W128 replaces Y107.

The possibility of DHDPS activity was investigated through a DHDPS assay as well and DSF to see if substrate binding took place. The kinetic assaying didn't show any DHDPS activity taking place for tHBPHA, this could be due to protein being catalytically inactive. DSF showed both proteins were binding to pyruvate which is expected due to it being a known substrate involved in the tHBPHA reaction.

3. Chapter three: What's the Hype about HypD?

3.1. Introduction into the Chapter

The protein of interest in this chapter is another members of the DHDPS/NAL subfamily, HypD (Δ^1-Pyrroline-4-hydroxy-2-carboxylate Deaminase or Pyr4H2C). This is a protein involved is involved in the degradation of L-hydroxyproline in bacteria. Four proteins will be structurally characterised within this chapter in order to determine the functional oligomeric state with a crystal structure being produced for *M. lupine* HypD.

3.1.1. Δ^1-Pyrroline-4-hydroxy-2-carboxylate Deaminase (HypD)

3.1.1.1. The HypD Enzymatic Pathway

L-Hydroxyproline mainly exists in collagen and within the cell wall of both plants and algae (Shibasaki, Mori et al. 1999). Its metabolism has been widely studied for mammals, involving a stepwise metabolism pathway that involves use of four different mitochondrial enzymes to produce the final products of pyruvate and glyocylate from a reaction with HOG aldolase (EC 4.1.3.16) (another member of the DHDPS/NAL subfamily (Riedel, Johnson et al. 2011). A similar pathway also takes place for L-proline, which results in the production of α-ketogltarate.

L-Hydroxyproline acts as a non-useable amino acid for a majority of microorganisms but few can metabolise and survive on L-hydroxyproline as a carbon source (eg *Pseudomonas putida)*(Gryder and Adams 1969). The microbial pathway is different from than found in mammals, first involving L-hydroxyproline epimerization into D-hydroxyproline and subsequent oxidation into Δ1-pyrroline-4-hydroxy-2-carboxylate (figure 21). This product then undergoes spontaneous hydrolysation into 4-hydroxy-2-oxo-5-aminovalerate. This is the substrate for HypD (Pyr4H2C). HypD then performs a deamination to form α-

ketoglutaricsemialdehyde, which then undergoes a $NADP^+$ dependent dehydrogenation into the final product α-ketoglutarate, a molecule involved as an intermediate in the Krebs cycle (Watanabe, Kodaki et al. 2006, Watanabe, Yamada et al. 2007).

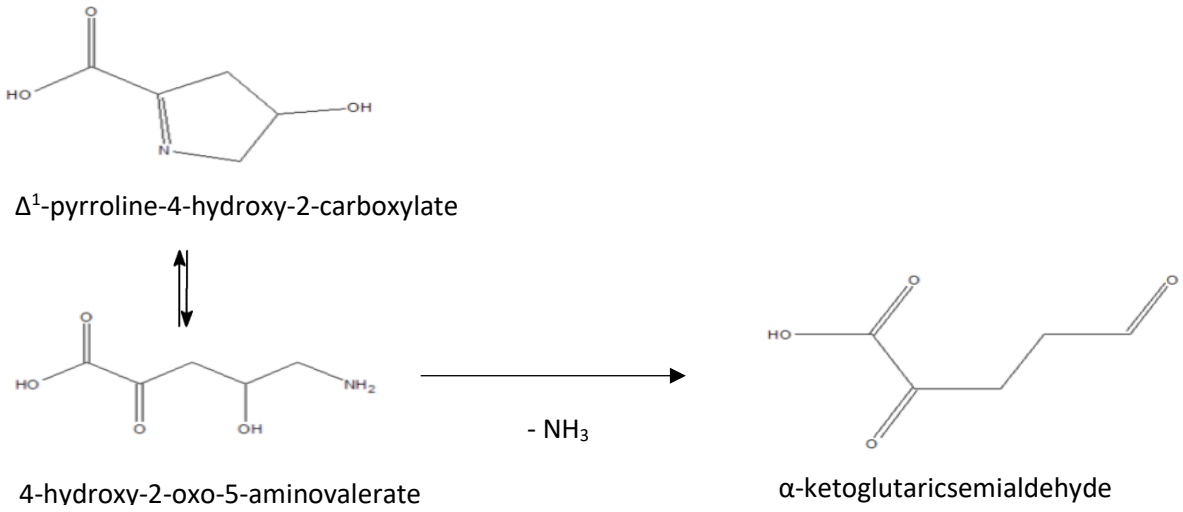


Figure 21

The reaction of HypD proteins is shown from the metabolic pathway of L-hydroxyproline. Δ^1-pyrroline-4-hydroxy-2-carboxylate is produced from the pathway and is undergoes spontainously Hydrolysation into 4-hydroxy-2-oxo-5-aminovalerate. HypD catalyses the deamination into α-ketoglutaricsemialdehyde (Watanabe, Kodaki et al. 2006, Watanabe, Yamada et al. 2007).

3.1.1.2. Past Studies Investigating HypD

One study investigated HypD from *P. putida* and *P. aeruginosai* and found that the two HypD proteins from these bacteria belong to the DHDPS/NAL subfamily of aldolase proteins. They examined the DHDPS activity and found that pyruvate was acting as an inhibitor (Watanabe, Morimoto et al. 2012). Because HypD shares the conserved lysine, which is involved in binding of pyruvate in most DHDHPS like proteins, its presence could act to bind pyruvate in a form of competitive inhibition in place of its substrate.

Another paper investigated a HypD protein from *S. meliloti* and have produced a crystal structure. They were able to structurally characterise a protein which was found to be within the DHDPS/NAL subfamily of proteins and exhibited the $(\alpha/\beta)_8$ TIM barrel fold. They reported that "The asymmetric unit of the solved structure is two monomers, and putative protein-protein interfaces suggested the formation of a homotetramer" (Chen, White et al. 2016)

3.1.2. Similarities to the DHDPS/NAL Subfamily

3.1.2.1. Sequence alignment of Related Proteins.

As with the previous chapter: the four HypD proteins examined within this thesis were subject to sequence alignment with other members of the DHDPS/NAL subfamily of proteins (figure 22). The sequence alignment is between: *B. melitensis* HypD*, A. tumefaciens* HypD*, S. meliloti* HypD*, M. lupini* HypD, *E. coli* DHDPS, *E. coli* NAL, *P. aeruginosa* DHDPS and *H. sapien* HOGA. Here we see some differences between the conserved residues of the active site between the reference proteins for this subfamily (*E. coli* DHDPS/NAL) the target proteins (the four HypD) and human HOGA protein (which reacts in a similar pathway from mammals (Riedel, Johnson et al. 2011))

The highly conserved lysine is present in the active site of the HypD proteins along with tyrosine which is also conserved through all members shown. These are the only two conserved amino acids within the active site for the HypD proteins. This suggests the catalytic mechanism could be significantly different to the other DHDPS like proteins. This is because there is no residue which corresponds to the Y107 in *E. coli* DHDPS which protrudes from the opposite monomer. This suggests that the active site architecture could be significantly different to other members of this subfamily. There are also no residues corresponding to T44 from *E. coli* DHDPS in the HypD structure but serine is conserved within this position for the other proteins in this alignment. This serine is likely interact in a similar way to threonine in DHDPS.

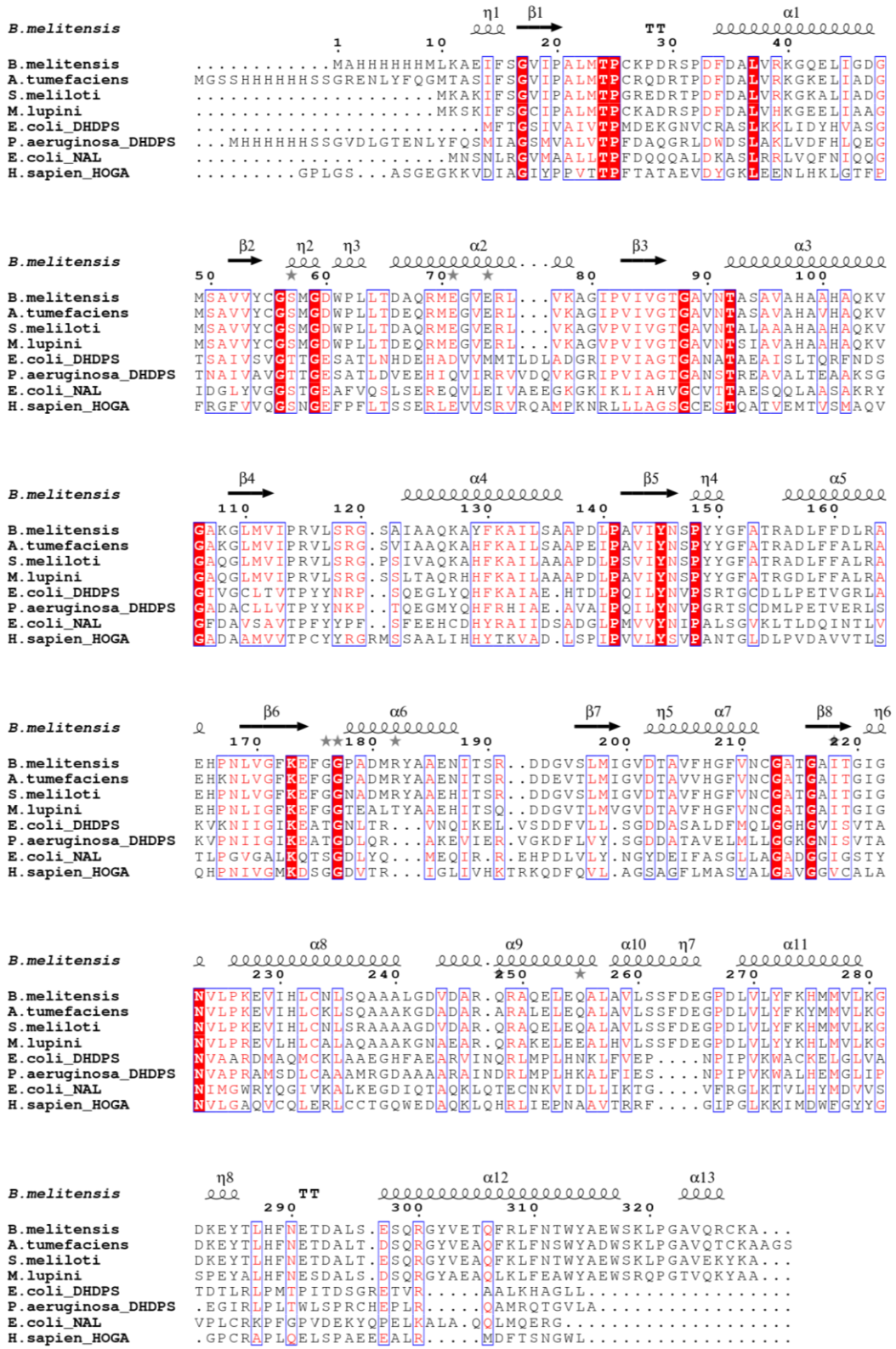


Figure 22

This figure shows the sequence alignments of some subfamily against the four HypD proteins. This figure shows: *B. melitensis* HypD*, A. tumefaciens* HypD*, S. meliloti* HypD*, M. lupini* HypD, *E. coli* DHDPS, *E .coli* NAL, *P. aeruginosa* DHDPS and H. sapien HOGA. This alignment is displayed using escript3.0 software (Robert and Gouet 2014) with secondary structure of *B. melitensis* HypD being shown along the top. Residues that are conserved throughout are in bold red while similar amino acids are in a blue box. There are a number of conserved residues throughout, in particular lysine and tyrosine which align to the active sites of all members in this alignment.

(Watanabe et al) investigated the HypD proteins and determined that HypD might be in the DHDPS/NAL subfamily. Due to the conserved lysine and tyrosine residues this seemed to be the case, although HypD proteins don't react with pyruvate which is a norm for members of this subfamily (Watanabe, Morimoto et al. 2012). HypD seem to share a common ancestor to other members to the DHDPS/NAL subfamily but have evolved in a different way as shown with the likely difference in catalytic mechanism.

3.1.2.2. Crystal structure comparison of HypD to *E. coli* DHDPS

The crystal structure of *S. meliloti* HypD (5CZJ) is shown in its reported tetrameric orientation (figure 23 a)). This protein exhibits the standard TIM barrel fold that is found within the DHDPS/NAL subfamily. The interesting point is that the two dimers are slightly angled, resulting in formation of a gap in the structure. When compared to the tetrameric structure of *E. coli* DHDPS (figure 23 b)), the presence of the gap shown with *S. meliloti* HypD is absent. This gap could be large enough to allow for a third dimer to fit, resulting in a hexameric protein. Figure 23 c) shows the *A. tumefaciens* HypD (2HMC) in such a hexameric orientation. The third dimer is present at the rear of the protein where the gap in the *S. meliloti* HypD resided. This shows possibility for HypD to exist in a hexameric orientation.

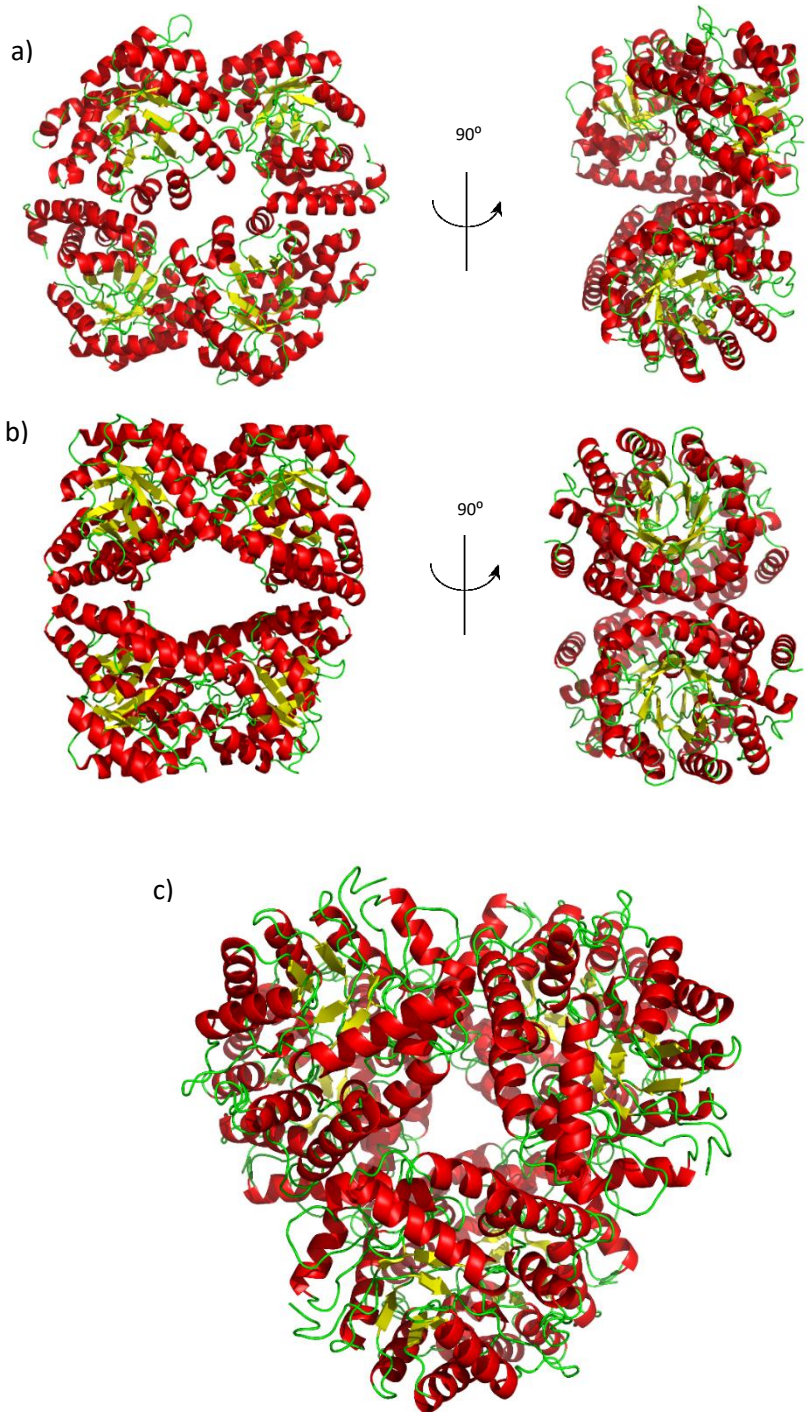


Figure 23

The tetrameric structure of *S. meliloti* HypD (5CZJ) as proposed by (chen et al) (a) compared to *E. coli* DHDPS (1YXC)(b). The α-helices are coloured red and β-sheets are coloured in yellow both proteins show a face on view (left) and a 90° rotation to show the side of the tetramer. *S. meliloti* HypD shows a large space when rotated compared to the lack of any space in DHDPS. The hexameric structure of *A. tumefaciens* HypD (2HMC) is shown (c). This structure shows the standard TIM barrel fold of the DHDPS/NAL subfamily but is present in a hexameric orientation as a trimer of dimers.

3.1.3. Aims for this chapter

The aim of this chapter is to investigate the structural characteristics of HypD. This will revolve around the determination of the oligomeric state of HypD through use of AUC and SAXS techniques to investigate the HypD from the previously characterised *A. tumefaciens, S. meliloti* and *B. melitensis* HypD. These proteins have been previously characterised, but the determined oligomeric states have been inconsistent due to different groups reporting HypD as a tetramer or hexamer. Finally a previously uncharacterised HypD (*M. lupine*) will be structurally characterised in this chapter. The possibility for DHDPS activity will be briefly investigated though a DHDPS assay

3.2. Results

3.2.1. Purification of the HypD proteins

The protein purified through his tag chromatography, with overnight TEV incubation followed by SEC chromatography to remove any TEV, his tag or aggregated protein. All HypD proteins had a high yield. *S. meliloti* and *B. melitensis* HypD produced about 30-40 mg/ml of protein from 1. 6L of media. *A. tumefaciens* and *M. lupine* HypD produced about 20-30 mg/ml from 1.6 L of media. The yields were comparable to the yields of *E. coli* DHDPS, *E. coli* NAL and *P. fluorescens* tHBPHA purified previously.

3.2.2. Analytical ultracentrifugation of HypD proteins

The first aim was to determine the oligomeric states of the four HypD proteins, which was first investigated through use of AUC. The four HypD proteins were analysed using *E.* coli DHDPS as a reference for a tetrameric protein. This aims to determine if the proteins are in a tetrameric or hexameric state. All HypD proteins were at a concentration of 0.5 mg/ml and

E. coli DHDPS from the previous chapter (was used as a tetrameric reference) was at 0.5 mg/ml. Each protein was showing a reading of about 0.5 at 280 nm.

3.2.2.1. Sedimentation of HypD proteins compared to E. coli DHDPS

The four HypD proteins show a significant difference in sedimentation coefficients. The previously described sedimentation of the tetrameric DHDPS was 7.25 S. The results for the AUC for the HypD proteins are shown with figure 24. The HypD proteins produced sedimentation values of: *A. tumefaciens* 9.6 S, *S. meliloti* 9.2 S, *B. melitensis* 9.4 S, *M. lupine* 9.4 S. The sedimentation of the HypD proteins is compared with that of *E. coli* DHDPS. This shows that the sedimentation of HypD is significantly different to that of *E. coli* DHDPS. Because the large difference between the tetrameric DHDPS and the HypD proteins it suggests that the HypD proteins are far larger than DHDPS. There are no other species present within the solution, suggesting that HypD is only in one state within solution.

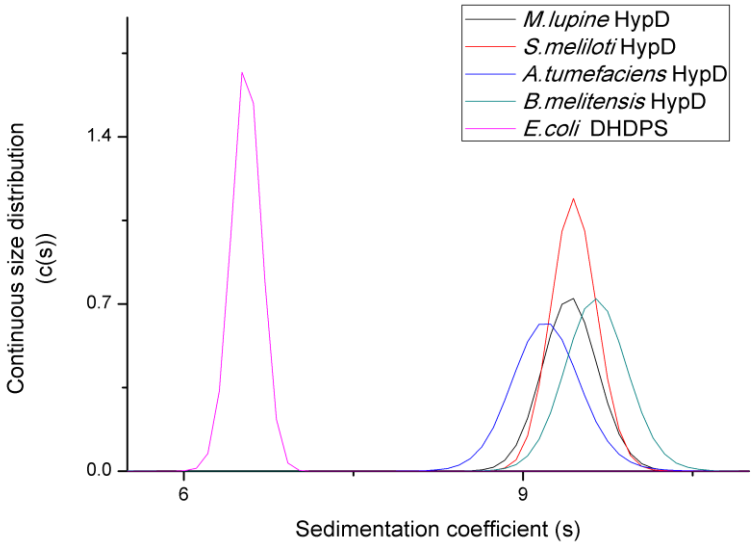


Figure 24

The sedimentation coefficients of the four HypD proteins were analysed using AUC sedimentation velocity. The sedimentation values were: *B. melitensis* at 9.6 S, *A. tumefaciens* at 9.2 *S. meliloti* at 9.4 S and *M. lupine* at 9.4 S. All four HypD proteins show a consistent sedimentation with no obvious other species present within solution as well as a significant difference in sedimentation when competed to DHDPS

3.2.2.2. Integrated Mass of HypD Proteins

The molecular mass of the four HypD proteins was also determined using AUC sedimentation velocity. As with the previous tHBPHA chapter, after determining the molecular mass from AUC the oligomeric state can be determined based on the molecular mass of the protein sequence. Because there is only one species present from the HypD proteins, calculation of molecular

B. melitensis HypD shows the presence of a hexameric protein. It shows a theoretical monomeric weight of about 36.6 kDa based on the sequence. This produces a theoretical tetrameric weight of 142.4 kDa and a theoretical hexameric weight of 213.7 kDa. AUC estimated the molecular mass to be about 205 kDa. This value is significantly closer to that of a hexameric protein than a tetrameric one showing that *B. melitensis* HypD is in a hexameric orientation.

A. tumefaciens HypD has a sequence weight of 34.4 kDa which results in a tetrameric weight of 137.8 kDa or hexameric weight of 206.7 kDa. AUC estimated the molecular mass to be about 202.4 kDa. This weight is consistent with that of a hexameric weight, not a tetrameric weight.

S. meliloti HypD has a sequence weight of 34.4 kDa, resulting in a theoretical tetrameric weight of 137.8 kDa and a theoretical hexameric weight of 203.7 kDa. AUC estimated that the molecular mass to be about 198.3 kDa. This weight is consistent with that of a hexameric protein as opposed to that of a tetrameric one. This contrasts what *(Chen, White et al. 2016)* reported when they discovered a tetrameric HypD from *S. meliloti (Chen, White et al. 2016)*.

Finally, *M. lupine* HypD has a sequence weight of 34.3 kDa, resulting in a theoretical tetrameric weight of 137.3 kDa or a theoretical hexameric weight of 206 kDa. The calculated molecular mass from AUC was 209.6 kDa. This is, again, consistent with that of a hexameric protein and not a tetrameric one.

The calculated molecular mass from AUC are all consistent with hexameric proteins, showing that HypD proteins are hexameric in solution. Because there was no other species

present in the AUC profile, the result can't be skewed and is only based on the present protein.

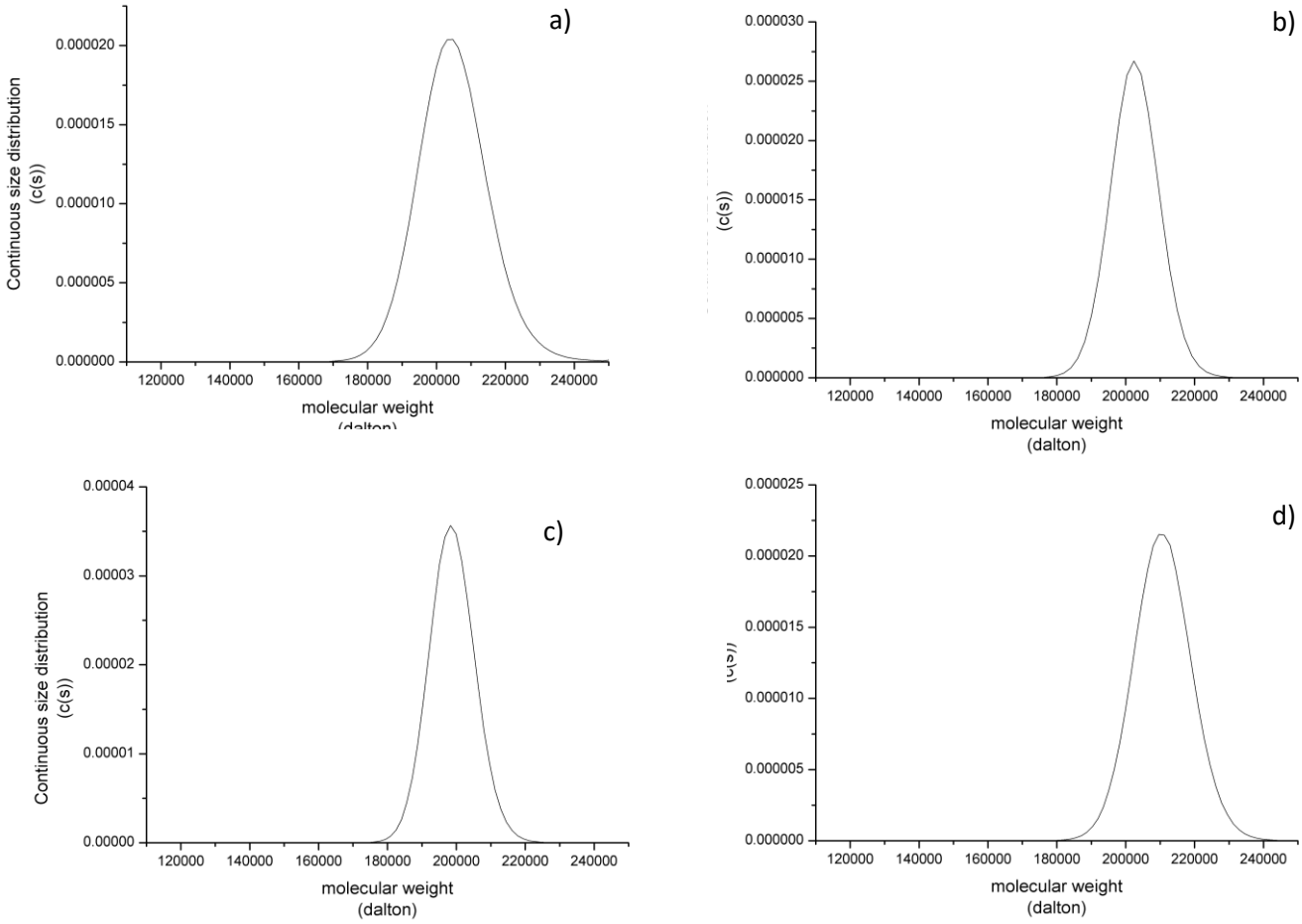


Figure 25

The molecular weights of the four HypD proteins were calculated using AUC sedimentation velocity. The calculated molecular mass was: *B. melitensis* at 205kDa (a*), A. tumefaciens* at 202.4kDa (b), *S. meliloti* at 202.4kDa (c) and *M. lupine* of 209.6kDa (d). All of the HypD proteins showed molecular weight values which are consistent with hexameric proteins.

3.2.3. Small angle x-ray scattering

In order to investigate the solution structures of the four HypD proteins, SAXS is used. All four HypD proteins are analysed using SAXS and the intensity data is plotted for *B. melitensis*, *A. tumefaciens*, *S. meliloti* and *M. lupine* (figure 26). These intensity plots show a high similarity, showing that the four HypD proteins share a similar form within solution.

3.2.3.1. SAXS Intensity data

The four intensity plots were normalised and aligned along with *E. coli* DHDPS. All four HypD protein share a very similar plot but contrast in shape to DHDPS. This suggests that the four HypD proteins aren't in a tetrameric orientation and show a different shape to DHDPS.

B. melitensis HypD shares a weight consistent with a hexameric protein. It produced a porod volume of 303427, with an I_0 value of 0.12 and an R_g of 37.79. The molecular weight of 4MPQ was probed with SAXS. Thig gave an estimated molecular weight of 205.7 kDa. The theoretical molecular weight of *B. melitensis* HypD, based of the original sequence, is a tetrameric weight of 142.5 kDa and a hexameric weight of 213.7 kDa. The closer molecular weight is closer to the theoretical hexameric weight suggesting that a hexameric orientation is present.

A. tumefaciens HypD is consistent with a hexameric weight. It produced a porod volume of 352273, with an I_0 value of 0.17 and an R_g of 40.52. The theoretical molecular weight of the *A. tumefaciens* HypD tetramer is 148.3 kDa with a hexameric weight of 222.5 kDa. SAXS has estimated that he total molecular weight of about 237.7 kDa. This weight is about 15 kDa more than the theoretical hexameric weight and is significantly larger than the tetrameric weight.

S. meliloti HypD is consistent with the weight of a hexameric protein. It produced a porod volume of 293859, with an I_0 value of 0.18 and an R_g of 38.31. The theoretical molecular weight of the *S. meliloti* HypD tetramer is 137.8 kDa with a hexameric weight of 203.7 kDa. SAXS calculated that the weight of the solution structure was 211.6 kDa. This weight is

comparable to that of a hexameric protein, contrasting to a tetrameric protein as reported previously (Chen, White et al. 2016).

The uncharacterised *M. lupine* HypD is consistent with the weight of a hexameric protein. It produced a porod volume of 303452, with an I_0 value of 0.078 and an R_g of 39. The theoretical molecular weight of the *M. lupine* HypD monomer is 34.3kDa, resulting in a tetrameric weight of 137.3 kDa or a hexameric weight of 206 kDa. The SAXS experiment has estimated that the total molecular weight of the solution structure is 198.6 kDa. Which results in a monomeric weight of 33.1 kDa. This value is comparable, again, to the theoretical hexameric structure, not the tetrameric structure.

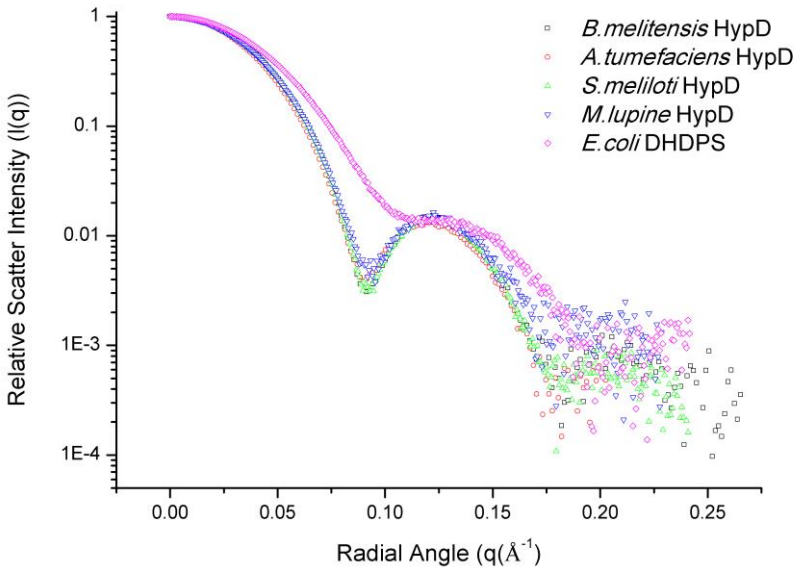


Figure 26

The intensity plots produced from SAXS results. They show the intensity plots of *E. coli* DHDPS and four HypD proteins: *B. melitensis*, *A. tumefaciens*, *S. meliloti*, *M. lupine*. The intensity plots are aligned with DHDPS acting as a tetrameric reference. This figure showing that all four proteins are in a similar orientation within solution but different to the tetrameric reference.

3.2.3.2. p(r) Plots Showing a Difference between HypD and DHDPS

With the exception of *A. tumefaciens* HypD, the HypD proteins show very similar molecular weights when compared to the theoretical hexameric weights. The four protein p(r) plots are aligned and normalised along with the p(r) plot of *E. coli* DHDPS (figure 27). All four protein show a similar p(r) distribution which is different from that of the tetrameric reference DHDPS, showing that the four HypD show a different presence in solution than *E. coli* DHDPS. This difference is likely due to the HypD proteins existing in a different conformation to DHDPS, due to the differences between the tetrameric DHDPS and hexameric HypD.

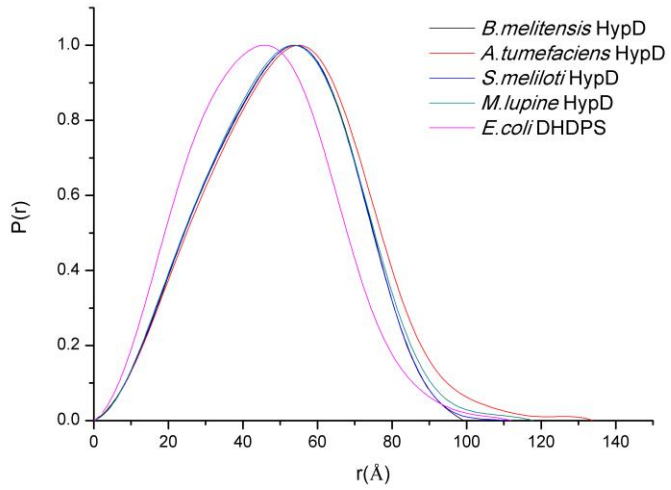


Figure 27

The p(r) data was calculated, using PRIMUS (Konarev, Volkov et al. 2003), for the four HypD proteins are plotted. This figure shows: *B. melitensis* HypD, *A. tumefaciens* HypD, *S. meliloti* HypD, *M. lupine* HypD. These plots are then normalised and aligned along with *E. coli* DHDPS as a tetrameric reference.

3.2.3.3. Kratky plots showing that HypD is a globular, folded protein.

A Kratky plot was used to probe for protein folding. All four proteins show a similar Kratky plot with a single, uniform peak between 0-0.15 Å and scatter past 0.25 Å. This shows that all present proteins are folded globular proteins within solution.

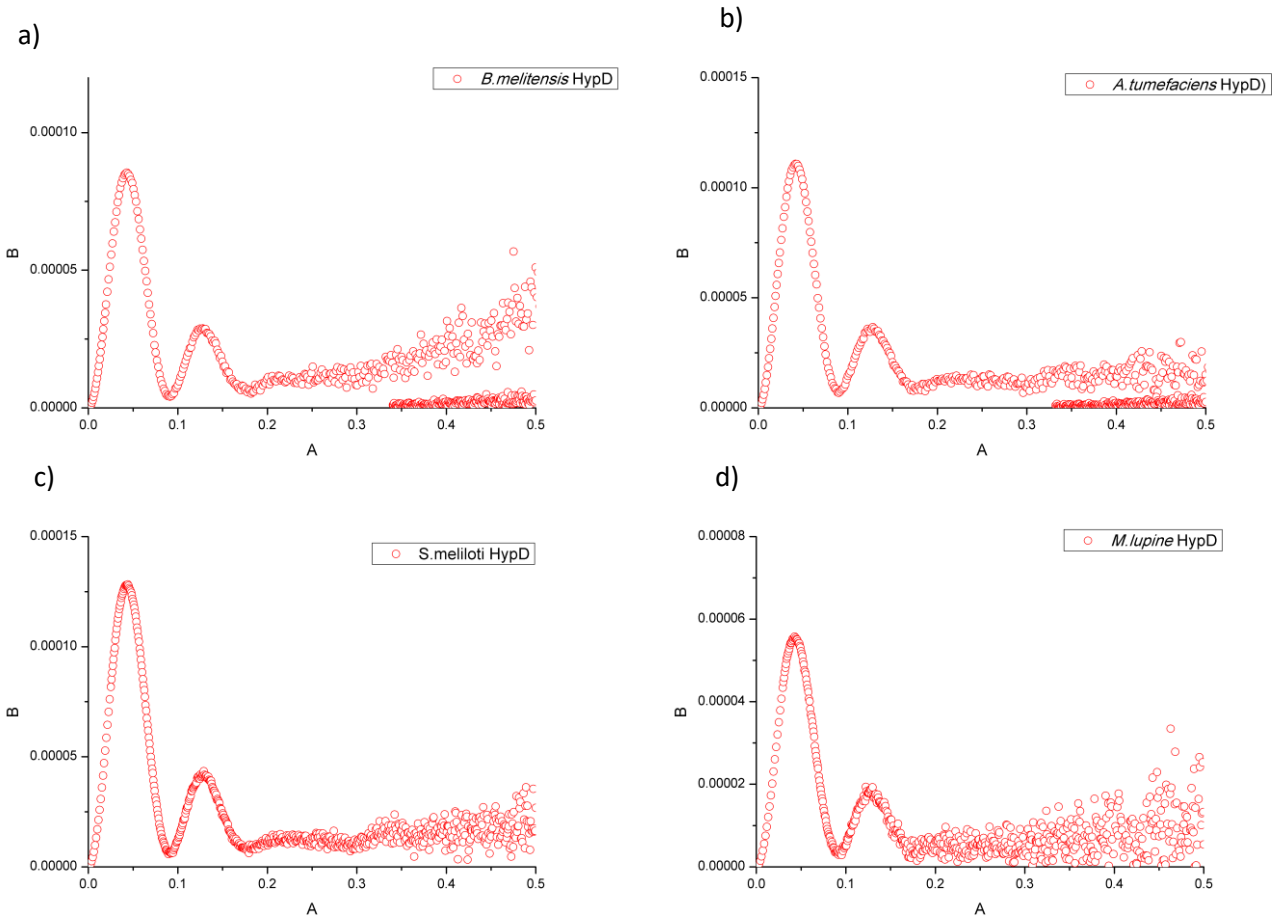


Figure 28

Four HypD shown with Kratky plots. The four plots show: B. melitensis HypD (a), A. tumefaciens HypD (b), S. meliloti HypD (c) and M. lupine HypD (d). All four proteins are folded within solution.

3.2.4. X-ray crystallographic study of *M. lupine* HypD

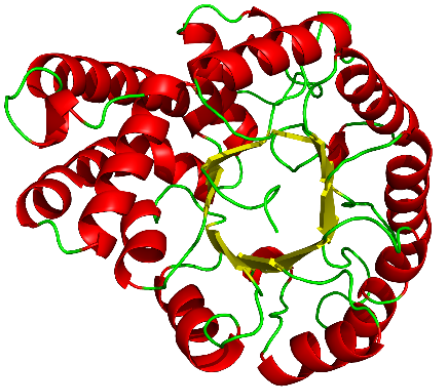
Three of the four HypD proteins examined within this thesis have solved crystal structures. *B. melitensis*(4MPQ), *A. tumefaciens*(2hmc), *S. meliloti*(5CZJ) all have crystal structures on the pdb, this project was able to produce a *M. lupine* crystal structure to 2.9 Å. The protein was crystallised within the PACT suite within the numerous conditions, E4 (0.2 M Potassium thiocyanate, 20% (w/v) PEG 3350). The structure was solved using *B. melitensis* HypD as a molecular replacement model with the final structure showed an R-work value of 0.28 and an R-free value of 0.34.

3.2.4.1. M. lupine HypD is A Hexameric Protein

The monomeric structure of *M. lupine* HypD shows the characteristic TIM barrel fold similar to that of other members of the DHDPS/NAL subfamily (figure 29, a)). Eight β-sheets are present within the protein forming a pore. Eight α-helices surround the inner barrel with a further three helices present on one side.

The crystal structure is shown in a hexameric orientation. (Figure 29, b)) Shows the hexameric orientation of *M. lupine* HypD. One way to describe the structure would be a comparison to a description of DHDPS. If DHDPS exists as a dimer of dimers then the observation of the HypD structure would be a trimer of dimers. The dimers also form a slight twist where the dimer-dimer interface meets, resulting in an angle which makes the dimers (curve). This contrasts to DHDPS which is more flat. Two monomers form an interface in a similar way to DHDPS and other members of the sub family. When two dimers come together within the crystal structure it somewhat resembles other DHDPS/NAL subfamily members but the presence of a large gap is also present as was shown in (figure 23, a)). This gap is large enough for a third dimer to sit, resulting in a hexameric protein.

a)



b)

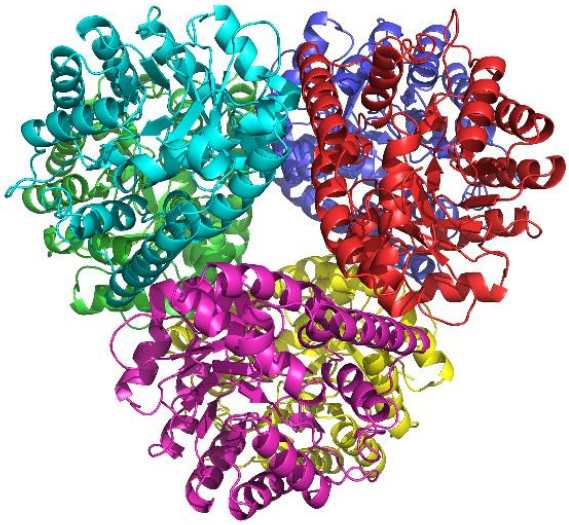


Figure 29

The crystal structure for *M. lupine* solved to 2.9 Å. The monomeric structure is shown (a) with α-helices coloured red, β-sheets in yellow and the remaining structure in green. This protein is shown to exist in a hexameric orientation and shown with each chain in a different colour (b).

3.2.4.2. Comparison of theoretical crystal scatter to solution scatter

The because all four proteins now have crystal structures produced form them, *B. melitensis* (4MPQ), *A. tumefaciens* (2hmc), *S. meliloti* (5CZJ) and *M. lupine*, the theoretical scatter generated from the crystal structures can be compared to the solution structure scatter from SAXS.

The comparison between a tetrameric HypD protein (as was described in (Chen, White et al. 2016) and a hexameric HypD is shown with Figure (figure 30). For *B. melitensis* (4MPQ), the tetrameric structure produced a chi^2 value of 31.081 compared to a chi^2 value of 0.535 for the hexameric structure. *A. tumefaciens* HypD (2HMC) produced a chi^2 value of 93.099 for the tetrameric structure while producing a chi^2 value of 2.559 for the hexameric structure. *S. meliloti* HypD (5CZJ) produced a chi^2 for the tetrameric structure of 54.237 while the hexameric structure produced a chi^2 value of 0.344. Finally, *M. lupine* HypD structure produced a chi^2 value of 34.237 for the tetrameric structure and 3.569 for the hexameric structure.

These chi^2 values strongly suggest that the HypD crystal structure is most correct when in a hexameric orientation. The significant differences between the tetrameric chi^2 values and the hexameric chi^2 values show that a hexameric HypD is the correct orientation for the crystal structures.

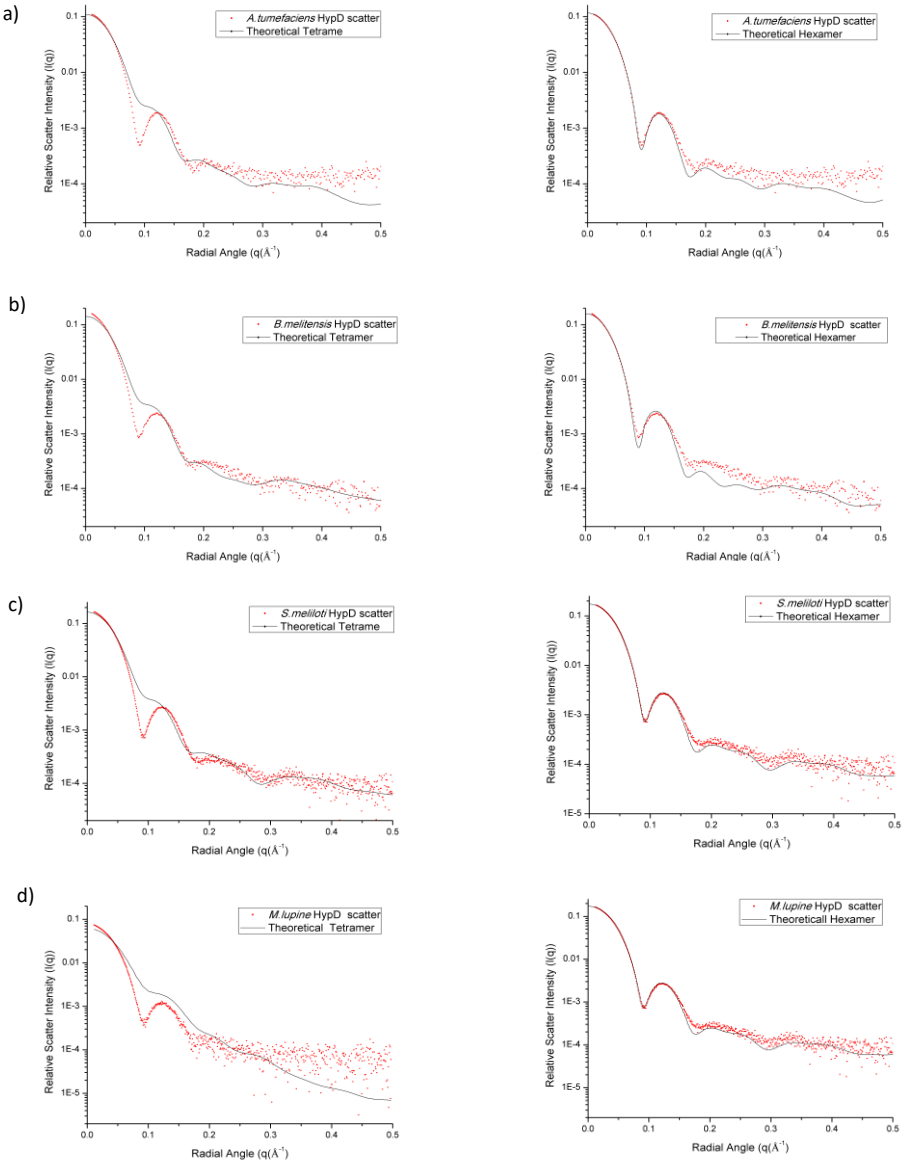


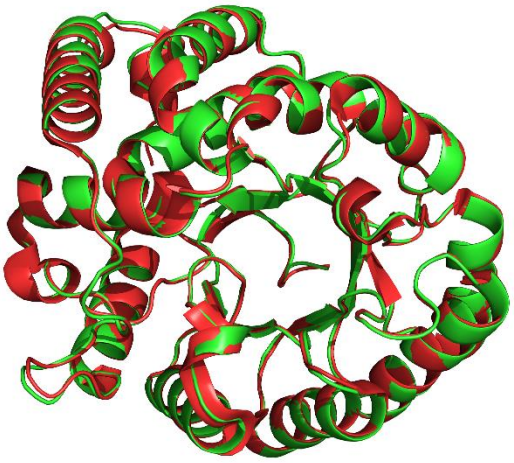
Figure 30

The theoretical scatter from the four HypD crystal structures was compared to the SAXS scatter to determine which oligomeric state is correct. The four HypD proteins: *B. melitensis* HypD(4MPQ)(a), *A. tumefaciens* HypD(2HMC) (b), *S. meliloti* HypD(5CZJ) (c) and *M. lupine* HypD(d). Each shows the theoretical scatter of a tetrameric crystal (left) compared to the theoretical scatter from a hexameric crystal (right).

3.2.4.3. Comparison of the crystal structures of *M. lupine* and *S. meliloti* HypD

The new crystal structure of *M. lupine* HypD and *S. meliloti* HypD are overlaid. Since *S. meliloti* HypD was previously reported as a tetramer, an overlay with the already confirmed hexameric *M. lupine* HypD would further confirm the presence of a hexamer through the architectural similarities. (figure 31, a)) Shows the monomeric alignment of the two proteins showing a high similarity. *M. lupine* HypD (green) shows a high similarity to the structure of *S. meliloti* HypD, this should be expected due to both being HypD proteins and sharing about an 81.5% sequence identity. The similarities are further confirmed when the two hexameric HypD are aligned as shown in (figure 31, b)). *M. lupine* HypD and *S. meliloti* HypD show a very similar hexameric orientation which only slight variation. This further confirms that *S. meliloti* HypD was miss-identified as a tetrameric protein

a)



b)

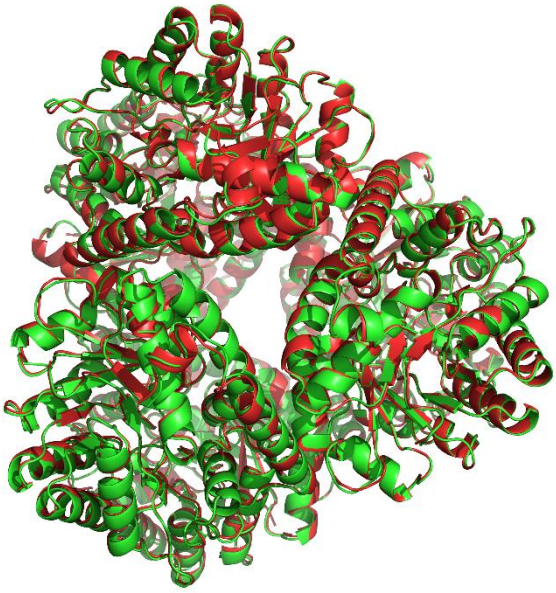


Figure 31

The monomeric crystal structures of the *S. meliloti* (red) and *M. lupine* (green) HypD are aligned to show the similarity between the two (a). The hexameric structure of *M. lupine* HypD is then aligned to the hexameric structure of *S. meliloti* HypD. This shows that *S. meliloti* HypD can align into a hexameric structure.

3.2.4.3.1. Active site of B. melitensis HypD and possible mechanism

The catalytic mechanism for HypD proteins is unknown due to lack of structures solved bound to substrate. Figure 31 shows the active site of *B. melitensis* HypD with the residues that are conserved with members of the DHDPS/NAL subfamily are shown in yellow (K164, and Y136) and residues surrounding these possible key amino acids in cyan. Me102 is also positioned near the conserved lysine and tyrosine. A possible interaction could occur. One point that contrasts with other members of the subfamily is that there doesn't seem to be any involvement from the opposite dimer within the active site. This suggests that each monomer can individually catalyse the reaction. S48 is in a similar position to serine found the NAL active site, this suggests that it could operate in a similar way. K 164 is almost certainly involved in Schiff base formation with the substrate, this is due to the similar mechanism present in all other members of the DHDPS/NAL subfamily. Because there hasn't been a structure solved with substrate residing within, the possible mechanism can only be speculation.

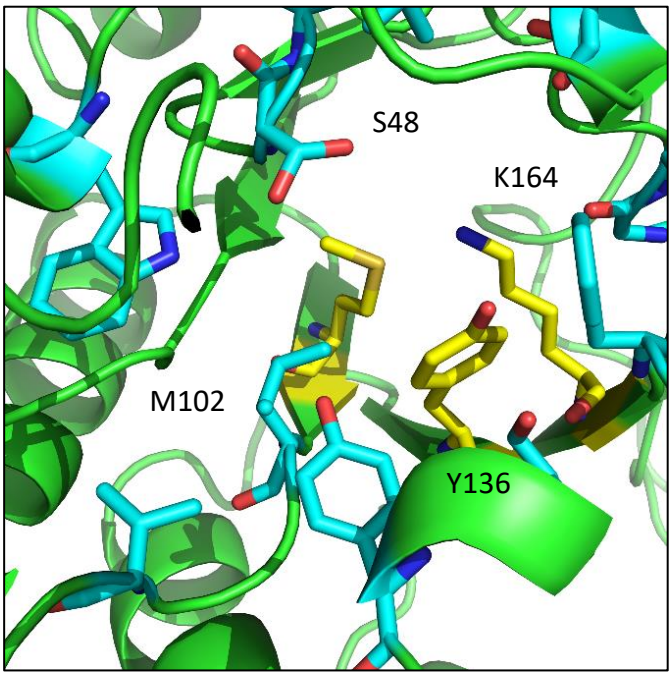


Figure 31

The active site of *B. melitensis* HypD (4MPQ) is examined. Amino acids which are identified through sequence alignments are coloured yellow with other residues possibly involved in catalysis are coloured cyan. The conserved residues of K173 and Y145 are present in HypD and other members of the DHDPS/NAL subfamily. S48 is in a similar position to that of the serine in the NAL active site.

3.2.4.3.2. Overlay of *B. melitensis* HypD with *E. coli* DHDPS

The monomers of DHDPS and HypD share a similar structure due to the TIM barrel fold but are slightly different. The monomers of *E. coli* DHDPS (red) and *B. melitensis* HypD (blue) are aligned (Figure 32). They both share the same fold but the DHDPS shows a 'tighter' structure when compared to HypD. The DHDPS structure's coils seem to be more compact than that of HypD, resulting in a slightly smaller monomer. The HypD monomer α-helices seem to be more spacious than that of DHDPS.

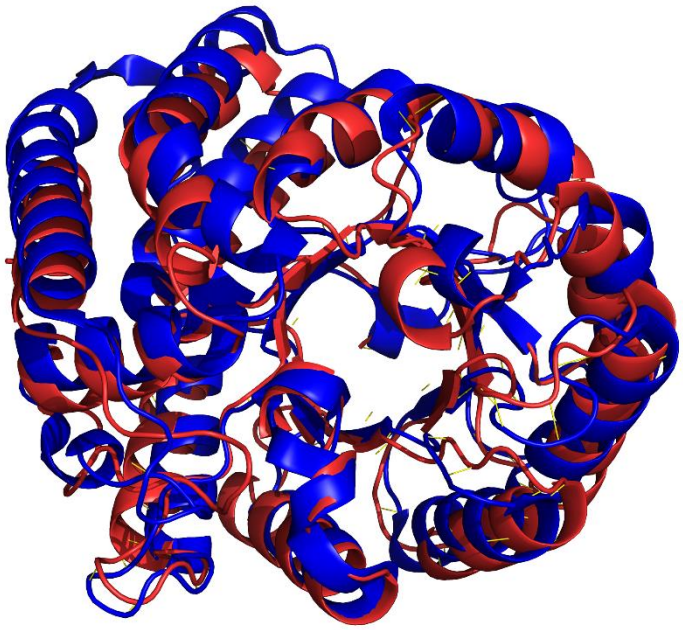


Figure 32

The alignment of *E. coli* DHDPS (red) and *B. melitensis* HypD (blue). Both share a similar fold, DHDPS seems to show a tighter arrangement of the secondary structure when compared to HypD.

3.2.5. Kinetic Assay for DHDPS Activity

S. meliloti and *M. lupine* were added in place of DHDPS in a DHDPS assay in order to determine if there was any DHDPS activity for both proteins. The kinetics showed that there was not any significant activity for either *S. meliloti* or *M. lupine*, concluding that, while both are within the same subfamily and share structural similarities, they don't exhibit the ability to catalyse a DHDPS reaction.

3.3. Discussion and Conclusion

3.3.1. HypD are hexameric proteins within solution

Initial AUC readings show that the way HypD sediments is significantly different to that of *E. coli* DHDPS. This is shown with (Figure 24) where the all four HypD proteins share a similar sedimentation coefficient ranging between 9.2 S and 9.6 S. When compared to the sedimentation vale of *E. coli* DHDPS or even the tHBPHA proteins of 7.1 S-7.3 S. This data suggests that the four HypD sediment with solution in a different way to the other tetrameric proteins within the DHDPS/NAL subfamily. One possibility for this is the presence of a hexameric protein. There is also a distinct lack of any secondary peaks on the sedimentation graph, showing that the possibility for two different oligomeric states based on concentration is unlikely. This difference is further seen when the AUC s(M) data is analysed, showing that the four HypD proteins investigated showing molecular mass of about 200 kDa, again consistent with hexameric proteins.

AUC has shown that the HypD proteins are unlikely to be tetrameric proteins, this is further confirmed through SAXS. SAXS estimated the solution molecular weights for all the HypD proteins and found that the weights were significantly larger than the theoretical tetrameric weights while being very similar to the theoretical hexameric weight. The intensity plots and

p(r) plots also show a noticeable difference when comparing the HypD proteins to their tetrameric subfamily members.

S. meliloti HypD is in a hexameric orientation which contrasts to what was reported previously by (Chen, White et al. 2016). The crystal structures of the Four HypD proteins are more comparable to the solution structures than their tetrameric orientations. This is shown with (figure 29) further showing the presence of the hexameric proteins.

The study that characterised *S. meliloti* HypD as a tetrameric protein likely only observed the monomer / dimer crystal structure and concluded the presence of a tetrameric protein without investigating this claim. When the *S. meliloti* HypD crystal structure (5CZJ) was compared to the new *M. lupine* HypD crystal structure, the similarities in the monomeric and hexameric orientations further show that *S. meliloti* HypD is hexameric.

Whether HypD proteins evolved into a hexameric protein from a tetrameric origin or was originally a hexamer is unknown. It is speculation that the HypD protein has possibly evolved from a tetrameric protein. Because it is within the DHDPS/NAL subfamily where a tetrameric orientation is a relative norm, there is some credibility in the possibility of a tetrameric origin, the catalytic significance of this hexameric protein could be investigated in future to demine the catalytic activity: firstly, of the HypD proteins towards their intended reaction and secondly, the effect of mutations resulting in a tetrameric or even dimeric/ monomeric protein.

3.3.2. HypD is a novel member of the DHDPS/NAL subfamily

HypD proteins are present in the DHDPS/NAL subfamily of proteins with the TIM barrel fold but is a novel member of this family. It is hexameric in a subfamily where the standard protein is in a tetrameric orientation. HypD also doesn't share the similar chemistry involving pyruvate. Most DHDPS are in a tetrameric orientation and mutations effecting the

oligomeric state negatively impact the enzymatic activity (Pearce, Dobson et al. 2008) while the hexameric HypD have added a unique touch to this subfamily.

The sequence alignment (figure 22) showed that there are certain differences between the active sties of the HypD proteins and other members of the DHDPS/NAL subfamily of proteins. The alignment showed that threonine was replaced with serine in HypD and that the second tyrosine in DHDPS (and tryptophan in tHBPHA) was replaced with valine and wasn't present in the active site. The lack of a similar amino acid in the position suggests there is no interaction between the opposing monomer. Showing that the catalytic mechanism of HypD should different to other subfamily members. There is still the highly conserved lysine present within the HypD active site which would result in binding and Schiff base formation with HypD's substrate. The serine residue is in a similar position to that of the serine in the NAL active site. It is positioned in a location where the substrate could bind within. The lack of an interaction between monomers is different from the other (known) subfamily members.

This speculation is due to the differences in the active site architecture where the monomer. As reported in (Watanabe et al), HypD proteins seem to share a common ancestor inspite not involving pyruvate (based in a phylogentetic tree)(Watanabe, Morimoto et al. 2012). HOG aldolaste is another member of the DHDPS/NAL subfamily which is involved in mamanlian L-Hydroxyproline metabolism but does involve a reactoin where pyruvate is the product(Watanabe, Kodaki et al. 2006, Watanabe, Yamada et al. 2007). The possibility of divergent prtoein evolutoin is a possiblitiy since with HypD retaining tyrosine and lysine within the active site but none of the other properties of DHDPS.

3.3.3. DHDPS activity from HypD?

As previously stated, the active site architecture of the HypD proteins suggest that DHDPS activity is unlikely. There is still the conserved lysine and tyrosine present in the active site but formation of a catalytic triad similar to DHDPS shouldn't be possible. It was reported by Watanabe et al), that pyruvate inhibited HypD, this could be due to pyruvate binding within the active site to the conserved lysine.

The possibility for HypD to show DHDPS activity was investigated through use of a DHDPS assay. This was done in order to determine if HypD (a member of the DHDPS/NAL subfamily) could share on a catalytic similarity with DHDPS and catalyse its reaction. If it does express DHDPS activity then the possibility of HypD acting as another DHDPS could lead to problems for drugs that specifically target DHDPS for inhibition. 5CZJ and MICRO were assayed to investigate this possibility. There was no change in the absorbance measurements when the HypD were added into the assay. This shows that, while being a member of the DHDPS/NAL subfamily of TIM barrel proteins, they don't exhibit DHDPS activity.

3.3.4. Conclusion

The four HypD proteins studied are seemingly in a hexameric orientation. AUC shows that the HypD sediment differently to a tetrameric reference in DHDPS. The SAXS intensity data contrasts between the HypD and DHDPS and estimated molecular weights of the HypD proteins are in line with hexameric proteins.

The crystal structure of *M. lupine* HypD was solved to 2.9 Å, showing a hexameric orientation. The crystal structures theoretical scatter matched the solution scatter from SAXS when a hexameric structure is used which further showed this fact. DHDPS activity was assayed and found no significant activity, showing that HypD can't act as a DHDPS in spite sequence similarities.

4. Chapter four: conclusions

The DHDPS/NAL subfamily of proteins are a good group to observe protein evolution due to all proteins sharing a commonly conserved TIM barrel fold and catalyse similar reactions. The Subfamily contains a number of proteins from different catalytic pathways but catalyse similar reactions. Of these, DHDPS and NAL are well characterised while others remain vaguely characterised. This thesis aims to characterise some of the more obscure members of this subfamily.

4.1. Aims

This project aims to structurally characterise two proteins from this subfamily, tHBPHA has previously been reported in a trimeric orientation (which is a direct contrast to other members of this subfamily) which could represent a novel member of this subfamily. HypD is another obscure member. Previously, HypD has been classed as a tetrameric protein or a hexameric protein. This thesis aims investigate the structural characteristics of these proteins as well as to determine the crystal structures in order to investigate the possibly different catalytic mechanisms. It also aims to investigate the possibility for these proteins to exhibit like DHDPS activity through use of a DHDPS kinetic assay.

4.2. tHBPHA is tetrameric

Through use of different analytical techniques, it was concluded that tHBPHA is a tetrameric protein. AUC showed that both *P. fluorescens* tHBPHA and *S. xenophagum* tHBPHA are consistant with tetrameric proteins. Their sedimentation coefficients are simmilar to that of the tetrameric reference *E. coli* DHDPS (7.0-7.3 S) as well as the calculated molecular mass being simmilar to the theoretical tetrameric value calculated from the sequence. With use of

SAXS this was further shown. The SAXS porod volumes for *P. fluorescens* tHBPHA further showed the presence of a tetrameric protein.

The crystal structure for *P. fluorescens* tHBPHA was solved to 2.2 Å. The solved structure showed the TIM barrel fold which conserved throughout the other members of this subfamily. The tetrameric orientation is similar to other DHDPS proteins. The active site is highly conserved with the only active site residue not conserved from *E. coli* DHDPS (W128 in tHBPHA) which replaces tyrosine. This residue could interact in a similar way to tyrosine with no crystal structures with substrate bound within the active site, the mechanism can only be speculation. To see if this similarity in the active site would result in DHDPS activity, *P. fluorescens* tHBPHA was assayed for DHDPS activity but there was no significant activity towards the DHDPS substrate.

4.3. HypD is hexameric

Inorder to determine which ologumeric state the HypD proteins, AUC sedimentation velocity was used. The HypD proteins showed a sedimentation coefficient signifigantly larger than the tetrameric reference of DHDPS, showing that HypD sediment different from tetrameric proteins. This was further shown when the molecular mass was calculated over 200 kDa for the four HypD proteins. This weight is consistant with the theoretical hexameric weight, showing the presence of hexameric proteins. Molecular weights calculated from SAXS porod volumes were also consistant with the hexameric proteins. when the theoreical scater from the crytsal structures were compare to the SAXS scatter, the result showed a higher correlation with a hexameric HypD than a tetrameric one.

The crystal structure for *M. lupine* HypD was producd and solved to 2.9Å. HypD show the conserved TIM barrel fold present within the subfamily. One interesting point is tha the HypD proteins don’t seem to show any interactions with residues within the active site and residues from the opposing monomer in the dimer interface. They lack a tyrosine (in DHDPS/NAL) or even a tryphtophan (in tHBPHA) that potrudes into the active site. This suggests that the mechanism for HypD is different from the mechanism for other subfamily members. HypD still retains the highly conserved lysine, while its reaction doesn not invlovle pyruvate, the mechanism should still involve the use of this residue to form a schiff base

intermeidate with the substrate. The possiblity for HypD to catalyse the DHDPS reaction was investigated, there was no signifigant activity from HypD.

4.4. Future research

Since both *P. fluorescens* tHBPHA and the four HypD proteins are able to form crystal structures, they can be used to investigate these proteins catalytic activity. How the catalytic mechanism would progress will be dependent of these substrate studies. How does the tryptophan in tHPBHA interact with its substrate? Is it similar to how tyrosine in DHDPS functions? This is also true for HypD, because there is no interaction between the two monomers in the dimer interface, the catalytic mechanism should follow a different mechanism. How these different mechanisms fit in with the other members of this subfamily is an unknown which could be investigated.

5. Chapter five : methods and experimental

5.1. Producing competent cells

E. coli BL-21 (DE3) pLysS cells are streaked out on agar plates containing chloramphenicol antibiotic and incubated at 37ºC overnight. A single colony was inoculated into a 5 ml LB (25 g/L) culture with 5 µl chloramphenicol and incubated overnight. 100 µl was taken and incubated into a 100 ml flask of LB and incubated at 37ºC on a shaker. When the A_{600} reached ~0.8 A.U., the cells are transferred into autoclaved centrifuge bottles and stored on ice for 15 min. The cells are harvested by spinning in a centrifuge at 4000 rpm, 4ºC for 10 min. The cell pellet is gently resuspended in 20 ml of autoclaved $CaCl_2$. The cells are again harvested by spinning in a centrifuge at 4000 rpm, 4ºC for 10 min. The cell pellet is gently resuspended in 4ml of autoclaved $CaCl_2$. 100 µl aliquots are stored in autoclaved Eppendorf tubes and flash frozen in liquid nitrogen for storage at -80ºC.

5.2. Transformations of desired proteins into BL 21 (DE3) cells

All plasmids are ordered from Epoch life science and are stored within the -80ºC freezer. A total of seven plasmids were ordered and transformed during this thesis. *E. coli* DHDPS and NAL plasmids were previously transformed into BL21 (DE3) pLysS cells and stored as a glycerol stock. All plasmids contained a His-tag and TEV cleavage region.

Agar plates used to grow transformed cells was prepared by mixing LB (15 g/L) and agar (10 g/L) and then autoclaved. The appropriate amount of chloramphenicol and kanamycin is added and the plates are poured onto trays in a sterile fume hood. Eppendorf's containing 100 µl competent cells are taken from the -80ºC freezer and thawed on ice. 1 µl of the desired plasmid is added and incubated on ice for 40 min. Heat shock each transformation tube by placing the bottom 1/2 of the tube into a 42 °C water bath for 45 min. 1 ml of LB media is added and the cells incubated at 37ºC on a shaking platform for 60 min. Transformed cells are harvested by spinning at 5000 rpm for 1 min, the remaining media is removed by

inverting the tube. The cells are resuspended in the remaining media. A control is prepared following the previous steps to ensure that the original cells cannot grow and only transformed cells can. The resuspended cells are then spread over the agar plates using a sterile plastic/glass spreader within a sterile fume hood. The cells are left for 10 min and then incubated at 37ºC overnight. If the transformation has succeeded, the plates will show colony growth.

5.3. Purification of proteins

tHBPHA was purified from *P. fluorescens and S. xenophagum* . HypD protein was purified from: *A. tumefaciens, S. meliloti, B. melitensis* and *M. lupine*. This project involved purifying a number of proteins in order to investigate structural details as well as for adding reference points using already known structures. *E. coli* DHDHPS and NAL were purified as a reference point for tetrameric proteins with *E.* coli DHDPS acting as a tetrameric reference for the tHBP-HA and HypD proteins.

One colony from the transformed cell plates is selected and transferred into a 5 ml culture of LB containing 10 µl of chloramphenicol and kanamycin. This is incubated at 37ºC on a shaking platform overnight. 1 ml of this culture will then be transferred into each large culture flask containing 800 ml of M9ZB media. These flasks will then be incubated at 37ºC on a shaking platform. The cultures will be monitored until the A_{600} reaches ~0.8A.U. 800 µl of 1 mM IPTG will then be added to each flask with the flasks then being incubated at 25ºC on a shaking platform overnight.

M9ZB media						
Chemical	Na_2HPO_4	NH4Cl	KH2PO4	NaCl	Bacto-tryptone	Yeast extract
concentration (g/L)	6	1	3	6	10	6

Table 1

The formula for M9ZB growth media for bacterial cell growth.

After incubation cells are poured into 400 ml centrifuge tubes and harvested by spinning at 8000 rpm for 8 min at 4ºC. The cells are resuspended in 10 ml (per centrifuge tube) of His-tag buffer a (50 mM Na_2HPO_4, 500 mM NaCl, 30 mM imidazole, pH 8). The resuspended cells are kept on ice and then sonicated using the UP200S (hielscher) for 20 min at 70% amplification with the cycle set to 0.5. The sonicated cells are then transferred into the small centrifuge tubes and spun at 15000 rpm for 15 min at 4ºC. The supernatant is carefully collected in a 50 ml falcon tube to reduce the amount of the pelleted cellular debris being collected. While and kept on ice the crude protein is filtered through a .22-µm filter to remove large cellular debris that could be present in the crude solution. All proteins are purified through use of His-tag chromatography using the Akta. The His-Trap HP, 5 x 5 ml (GE Life Science) was attached to the Akta and was equilibrated through washing initially with His-tag buffer B (50 mM Na_2HPO_4, 500 mM NaCl, 300 mM imidazole, pH 8) to remove anything that could be inside the column. His-tag buffer A is then flowed through until the UV reading stabilises. Crude protein is injected into the Akta along with His-tag buffer A and passed through the His-tag column. After all the crude has passed through, buffer is streamed until no crude remains within the column and only protein is attached within (as seen with the change in UV reading). His-tag buffer B is streamed through to elute the protein attached to the nickel within the column, the elution of protein is shown with a large peak in the UV readings depending on how much protein is eluted.

Histag buffer a			
Chemical	Na2HPO4	NaCl	imidazole
concentration (mM)	50	500	30
histag buffer b			
Chemical	Na2HPO4	NaCl	imidazole
concentration (mM)	50	500	500

Table 2

His-tag Buffer formula for His-tag buffer A and his-tag buffer B.

The protein purified from His-tag chromatography is then passed through a desalt column to remove excess salt and imidazole present. The desalt column is a low resolution Size exclusion column which separates proteins from the smaller salt molecules. Tris buffer (150 mM NaCl, 20 mM tris-HCl, pH 8) is streamed along with protein through the column. The protein will exit the column before the salt, result in a protein solution in the desired tris-buffer.

tris-HCL sotrage buffer		
Chemical	NaCl	tris-HCl
concentration (mM)	150	20

Table 3

Tris-HCl Storage buffer. Proteins are stored at -80^{o}C within this buffer as well as undergo experimentation within

The his-tag on the proteins is removed via TEV cleavage. Previously purified TEV protease will be incubated with a His-tagged protein post desalt. TEV protease is added at a ratio of 1 mg of TEV to 100 mg of his tagged protein. Proteins can be further purified (to remove aggregate and TEV) through use of a 24 mL Superdex™ 200 10/300 gel filtration column (GE Healthcare). This is a 24 ml column that will separate different proteins based on their size as they pass through the column. The flow rate is set to 0.4-0.5 mL/min in order to reduce the pressure below 1.2 mPa. Tetrameric protein will elute from the colum at ~14 ml while hexameric protein will elute at ~12 ml. Any aggregate will be eluted at 10-12 ml. TEV will elute after the protein due to its smaller size.

5.4. Purification of TEV protease

TEV protease is used to cleave the His-tag off proteins. TEV protease was purified from stock cells possessing resistances for chloramphenicol and kanamycin. The cells were grown in M9ZB media, incubated at 37ºC on a shaking platform until they reached an OD_{600} of 0.6-0.8 A.U. They are then induced with addition of 1 mM IPTG and incubated on a shaking platform overnight at 25ºC.

After incubation cells are poured into 400 ml centrifuge tubes and harvested by spinning at 8000 rpm for 8 min at 4ºC. The cells are resuspended in 10 ml (per centrifuge tube) of TEV buffer A (50 mM Na_2HPO_4, 500 mM NaCl, 10 mM imidazole, 5% glycerol, pH 8.0).The resuspended cells are kept on ice and then sonicated using the UP200S (hielscher) for 20min at 70% amplification with the cycle set to 0.5. The sonicated cells are then transferred into the small centrifuge tubes and spun at 15000 rpm for 15 min at 4ºC. TEV protease is then purified via his-tag chromatography using TEV buffer A and TEV buffer B (50 mM Na_2HPO_4, 500 mM NaCl, 500 mM imidazole, 5% glycerol, pH 8.0) in place of His-tag buffer A and his-tag buffer B respectively. After purification the protein is kept on ice and precipitated using ammonium sulphate to a concentration of 65% and is kept on ice for 15 min. The solution is spun at 6000 rpm at 4ºC for 10 min and the supernatant is discarded. The precipitated TEV is dissolved in TEV storage buffer (300 mM NaCl, 20 mM tris-HCl, 1 mM EDTA, 50% glycerol, pH 8.0). It is then distributed into 100 µl volumes in eppendorf tubes, flash frozen in liquid nitrogen and stored at -80ºC until needed.

TEV His-tag buffer A				
Chemical	Na2HPO4	NaCl	imidazole	glycerol
concentration (mM)	50	500	10	5%
TEV His-tag buffer B				
Chemical	Na2HPO4	NaCl	imidazole	glycerol
concentration (mM)	50	500	500	5%
TEV storage buffer				
Chemical	NACL	tris-HCl	EDTA	glycerol
concentration (mM)	300	20	1	50%

Table 4

The formula for three TEV buffers for use in TEV purification.

5.5. Analytical Ultracentrifugation (AUC)

AUC sedimentation velocity experiments are used to determine the molecular weight and oligomeric state of the protein samples. It involves using a pure protein sample that has passed through a 24 mL Superdex™ 200 10/300 gel filtration column (GE Healthcare) . The proteins were equilibrated to the tris buffer. 400 µl of tris buffer (used as a blank) and 380 µl of protein sample are loaded into the 12 mm double-sector cells with quartz windows in an eight hole (An-50) rotor. Once all samples are loaded, the rotor is inserted into the ProteomeLab XL-I (Beckman) analytical ultracentrifuge. The rotor is then equilibrated to 20ºC. The AUC was run at 50000 rpm at 20ºC until 100 scans 285 nm using a UV/Vis scanning optics until completion. These process was carried out under supervision of Dr Grant Pearce.

The concentrations of proteins analysed were: *P.fluorescens* tHBPHA 0.5 mg/ml, *S.xenophagum* tHBPHA 0.2 mg/ml, *E. coli* DHDPS 0.5mg/ml, *B.melitensis* HypD 0.4mg/ml, *A.tumefaciens* HypD 0.4 mg/ml, *S.meliloti* HypD 0.4 g/ml, *M.lupine* HypD 0.4 mg/ml. the final *P.fluorescens* tHBPHA was run at varying concentrations of : 0.1 mg/ml, 0.2 mg/ml and 0.4 mg/ml. Data provided from the AUC is analysed using the SEDFIT program(Schuck 2000).This program will run either a c(M) distribution to determine protein molecular mass or c(s) distribution to determine protein sedimentation coefficients. Buffer density is set to 0.99823 g/ml and viscosity to 0.01002 g/L for the standard tris buffer the proteins are stored in, as calculated sednterp (T.M. Laue 1992). Data analysed with this program will be represented as graph of either protein sedimentation or protein molecular weight.

5.6. Gel electrophoresis SDS-PAGE

Protein SDS-PAGE was executed using premade gels. Each sample will contain 2.5 µl of 4X dye, 5-10 µg of protein (~1-6 µl depending on initial protein concentration) and the remaining MiliQ H_2O to bring the total volume of the sample to at least 10µl. The samples are mixed be briefly using a benchtop centrifuge and incubated in a 70ºC water bath for 5

min. The premade gel is inserted into the gel electrophoresis apparatus with MES buffer being poured to fill the apparatus. The samples are then pipetted into the different lanes with the addition of 5 µl of Novex® Sharp Pre-Stained Protein Standard (Invitrogen). The gel is run at 166 V for 35 min in 1X NuPAGE® MES SDS Running Buffer. The gel is then removed from the apparatus and its casing removed. The gel is rinsed in H_2O and microwaved twice before being stained for 1 h in Simply Blue stain. The gel is then rinsed in H_2O and de-stained overnight.

5.7. Concentrating protein samples

Protein samples are concentrated using Vivaspin 6, 30,000 MWCO PES concentrators from Sartorius Stedim Biotech. Protein is added into the tube and then sun in the centrifuge at 4000 rpm at 4ºC for 4-5 min per spin. The concentration of protein is determined using NanoDrop ND 1000 spectrophotometer from Thermo Scientific. If the protein concentration was not at the desired point, the process is repeated until the desired concentration is achieved.

5.8. X-ray crystallography

Protein to be used for crystallography is purified and run through the size exclusion column. It is then concentrated using Vivaspin 6, 30,000 MWCO PES concentrators, aiming to have the final concentration at about 10 mg/ml. The protein crystal screens are obtained using the JCSG+ and PACT suite screens (Molecular Dimensions). 96 well plates are filled with the screen conditions using a 12 channel pipette. Both trays and proteins are taken to the UC chemistry department. With a 50 µl stock protein sample was prepared at 10 mg/ml. Use of Mosquito® Crystal automated liquid distributer (TTP Labtech) to distribute a 50 µl stock protein sample into 400 nL droplets. The trays are then sealed and carefully incubated in the desired conditions. tHBPHA was incubated at 20ºC while HypD crystals were incubated 8ºC and 25ºC) to facilitate crystal growth. Each tray is examined for crystal formation using light microscopy every second day and any crystal growth is recorded.

The crystal trays with suitable crystals are sent to the Australian Synchrotron in order to collect diffraction data. The crystals are placed under a light microscope and the crystal is examined. The seal around the condition is cut with a scalpel and the crystal is extracted into a loop. This is quickly placed into cryoprotectant (1:1 mixture of ethylene glycol and glycerol) and submerged into a bowl of liquid nitrogen. This loop with the crystal is kept within the liquid nitrogen until it is attached within the beamline room under a cryo stream. The crystal is positioned directly in the path of the beam line. One the crystal is correctly positioned, the beamline is activated and x-rays are shot through the crystal, producing a diffraction pattern. This diffraction patters is analysed using CCP4 to produce a MTZ file with the diffraction data (Bailey 1994, McCoy, Grosse-Kunstleve et al. 2007).

The structures in this thesis were solved through use of the Australian Synchrotron (Panjikar, Parthasarathy et al. 2005, Panjikar, Parthasarathy et al. 2009). THBPHA from *P. fluorescens* was solved using AUTO RIKSHAW. First the program Pirate (Cowtan 2000) was used to statistically modify the phase probability distributions to produce an improved version of the phase probability distributions. This was further modified using the program SHELXE for phase and density modifications (Sheldrick 2010). Because the sequence similarity of *P. fluorescens* tHBPHA was so low to other solved proteins, ordinary molecular replacement was unable to solve the structure. Through use of Buccaneer (Cowtan 2006, Cowtan 2008), a model was formed which was suitable for molecular replacement. Molecular replacement and subsequent structural refinement was carried out using PHENIX software (McCoy, Grosse-Kunstleve et al. 2007, Afonine, Grosse-Kunstleve et al. 2012). The crystal structure was solved to 2.2Å, with an r-work value from 0.3 to 0.18 and a change in the r-free values of 0.38 to 0.235. *M. lupine* had a high sequence similarity to other solved HypD structures and use of AUTO RICKSHAW was not needed. Again, the program PHENIX was used for the molecular replacement and refinements using the crystal structure of *B. melitensis* HypD (4MPQ) and a molecular replacement model (McCoy, Grosse-Kunstleve et al. 2007, Afonine, Grosse-Kunstleve et al. 2012). The crystal structure was solved to 2.9Å resolution producing an R-work value of 0.28 and an R-free value of 0.34.

5.9. Small Angle X-ray Scattering

Small Angle X-ray Scattering is a technique to probe the solution structures of proteins. All samples to be sent were run through a 24 mL Superdex™ 200 10/300 gel filtration column (GE Healthcare) to remove any aggregate. The samples are equilibrated into tris buffer and are concentrated above 5 mg/ml. Samples are sent to the Australian Synchrotron and data collected using the SAXS/WAXS beamline with the Pilatus 1M detector (170 mm x170 mm, effective pixel size, 172 x 172 m) as described (Keown, Griffin et al. 2013). The wavelength of the x-rays was 1.0332 Å. Each sample was flowed through 3 ml Superdex 200 5/150 gel filtration column (GE Healthcare), prior to passing through the beamline to ensure only protein will be present in the data and not any aggregate. As the samples are flowed through the beamline the different frames of the scatter are recorded.

The scatter is analysed using Scatterbrain analysis (Australian Synchrotron) using images summed from the buffer and the most concentrated part of the protein sample. This data is then analysed using Primus (Konarev, Volkov et al. 2003, Petoukhov, Franke et al. 2012) .The theoretical scatter from the crystal structures are compared to the scatter produced from SAXS using Crysol (Svergun, Barberato et al. 1995) molecular weights were determined from SAXS using the SAXS MoW2 web tool(Fischer, Neto et al. 2010).

5.10. Differential Scanning Fluorimetry

Protein is concentrated to 0.5 mg/ml in 20 mM tris buffer10 mM pyruvate and 10 mM lysine are used as substrates to test for binding of substrates related to DHDPS. Protein samples are mixed with SYPRO® Orange (Sigma) at a 1000 fold dilution of the dye. Four different conditions were employed per protein; one with only protein, two with either pyruvate or lysine present at 0.5 mg/ml concentrations and one condition with both pyruvate and lysine present. The thermal melt assay conducted by increasing the temperature to from 10ºC in

increments of 0.5ºC until a temperature of 100ºC was reached. This was run with the Applied Biosystems™ QuantStudio™ 3 Real-Time PCR System.

5.11. Assaying for DHDPS Activity

DHDPS activity was assayed using a coupled DHDPS assay. This involves coupling the activity of DHDPS (or tHBPHA/HypD) with the activity of DHDPR (the next protein in the DAP pathway), with the change in NADPH measured through spectrophotometer as described in (Renwick C. J. DOBSON 2004). The assay was performed using 1 ml cuvettes which contained: 100 mM Hepes, ~0.162 mM NADPH, 20µg *E. coli* DHDPR and a set concentration of pyruvate and ASA at 50 mM. This is because the assay is to determine if activity is present. The protein concentrations are at 2 mg/ml within the cuvettes. The reaction mix is incubated for 10 min at 30ºC before addition of substrate.

6. Chapter six : References

Afonine, P. V., R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart and P. D. Adams (2012). "Towards automated crystallographic structure refinement with phenix.refine." Acta Crystallographica Section D-Biological Crystallography **68**: 352-367.

Almagro-Moreno, S. and E. F. Boyd (2009). "Insights into the evolution of sialic acid catabolism among bacteria." Bmc Evolutionary Biology **9**: 16.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. H. Zhang, Z. Zhang, W. Miller and D. J. Lipman (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Research **25**(17): 3389-3402.

Altschul, S. F., J. C. Wootton, E. M. Gertz, R. Agarwala, A. Morgulis, A. A. Schaffer and Y. K. Yu (2005). "Protein database searches using compositionally adjusted substitution matrices." Febs Journal **272**(20): 5101-5109.

Bailey, S. (1994). "THE CCP4 SUITE - PROGRAMS FOR PROTEIN CRYSTALLOGRAPHY." Acta Crystallographica Section D-Biological Crystallography **50**: 760-763.

Blickling, S., H. G. Beisel, D. Bozic, J. Knablein, B. Laber and R. Huber (1997). "Structure of dihydrodipicolinate synthase of Nicotiana sylvestris reveals novel quaternary structure." Journal of Molecular Biology **274**(4): 608-621.

Blickling, S., C. Renner, B. Laber, H. D. Pohlenz, T. A. Holak and R. Huber (1997). "Reaction mechanism of Escherichia coli dihydrodipicolinate synthase investigated by X-ray crystallography and NMR spectroscopy." Biochemistry **36**(1): 24-33.

Bolt, A., A. Berry and A. Nelson (2008). "Directed evolution of aldolases for exploitation in synthetic organic chemistry." Archives of Biochemistry and Biophysics **474**(2): 318-330.

Bright, S. W. J. and P. R. Shewry (1983). "IMPROVEMENT OF PROTEIN-QUALITY IN CEREALS." Crc Critical Reviews in Plant Sciences **1**(1): 49-93.

Cerniglia, C. E. (1984). "MICROBIAL-METABOLISM OF POLYCYCLIC AROMATIC-HYDROCARBONS." Advances in Applied Microbiology **30**: 31-71.

Chen, S. Y., C. E. White, G. C. diCenzo, Y. Zhang, P. J. Stogios, A. Savchenko and T. M. Finan (2016). "L-Hydroxyproline and D-Proline Catabolism in Sinorhizobium meliloti." Journal of Bacteriology **198**(7): 1171-1181.

Copley, R. R. and P. Bork (2000). "Homology among (beta alpha)(8) barrels: Implications for the evolution of metabolic pathways." Journal of Molecular Biology **303**(4): 627-640.

Cowtan, K. (2000). "General quadratic functions in real and reciprocal space and their application to likelihood phasing." Acta Crystallographica Section D-Biological Crystallography **56**: 1612-1621.

Cowtan, K. (2006). "The Buccaneer software for automated model building. 1. Tracing protein chains." Acta Crystallographica Section D-Biological Crystallography **62**: 1002-1011.

Cowtan, K. (2008). "Fitting molecular fragments into electron density." Acta Crystallographica Section D-Biological Crystallography **64**: 83-89.

da Costa, T. P. S., A. C. Muscroft-Taylor, R. C. J. Dobson, S. R. A. Devenish, G. B. Jameson and J. A. Gerrard (2010). "How essential is the 'essential' active-site lysine in dihydrodipicolinate synthase?" Biochimie **92**(7): 837-845.

Dobson, R. C. J., M. D. W. Griffin, G. B. Jameson and J. A. Gerrard (2005). "The crystal structures of native and (S)-lysine-bound dihydrodipicolinate synthase from Escherichia coli with improved resolution show new features of biological significance." Acta Crystallographica Section D-Biological Crystallography **61**: 1116-1124.

Dobson, R. C. J., K. Valegard and J. A. Gerrard (2004). "The crystal structure of three site-directed mutants of Escherichia coli dihydrodipicolinate synthase: Further evidence for a catalytic triad." Journal of Molecular Biology **338**(2): 329-339.

Eaton, R. W. and P. J. Chapman (1992). "BACTERIAL METABOLISM OF NAPHTHALENE - CONSTRUCTION AND USE OF RECOMBINANT BACTERIA TO STUDY RING CLEAVAGE OF 1,2-DIHYDROXYNAPHTHALENE AND SUBSEQUENT REACTIONS." Journal of Bacteriology **174**(23): 7542-7554.
Ferrara, S., E. Mapelli, G. Sello and P. Di Gennaro (2011). "Characterization of the aldol condensation activity of the trans-o-hydroxybenzylidenepyruvate hydratase-aldolase (tHBP-HA) cloned from Pseudomonas fluorescens N3." Biochimica Et Biophysica Acta-Proteins and Proteomics **1814**(5): 622-629.
Fischer, H., M. D. Neto, H. B. Napolitano, I. Polikarpov and A. F. Craievich (2010). "Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale." Journal of Applied Crystallography **43**: 101-109.
Gefflaut, T., C. Blonski, J. Perie and M. Willson (1995). "Class I aldolases: Substrate specificity, mechanism, inhibitors and structural aspects." Progress in Biophysics & Molecular Biology **63**(3): 301-340.
Griffin, M. D. W., J. M. Billakanti, A. Wason, S. Keller, H. D. T. Mertens, S. C. Atkinson, R. C. J. Dobson, M. A. Perugini, J. A. Gerrard and F. G. Pearce (2012). "Characterisation of the First Enzymes Committed to Lysine Biosynthesis in Arabidopsis thaliana." Plos One **7**(7): 12.
Griffin, M. D. W., R. C. J. Dobson, F. G. Pearce, L. Antonio, A. E. Whitten, K. Liew, J. P. Mackay, J. Trewhella, G. B. Jameson, M. A. Perugini and J. A. Gerrard (2008). "Evolution of quaternary structure in a homotetrameric enzyme." Journal of Molecular Biology **380**(4): 691-703.
Gryder, R. M. and E. Adams (1969). "INDUCIBLE DEGRADATION OF HYDROXYPROLINE IN PSEUDOMONAS PUTIDA - PATHWAY REGULATION AND HYDROXYPROLINE UPTAKE." Journal of Bacteriology **97**(1): 292-&.
Hocker, B., S. Beismann-Driemeyer, S. Hettwer, A. Lustig and R. Sterner (2001). "Dissection of a (beta alpha)(8)-barrel enzyme into two folded halves." Nature Structural Biology **8**(1): 32-36.
Izard, T., M. C. Lawrence, R. L. Malby, G. G. Lilley and P. M. Colman (1994). "THE 3-DIMENSIONAL STRUCTURE OF N-ACETYLNEURAMINATE LYASE FROM ESCHERICHIA-COLI." Structure **2**(5): 361-369.
Jeong, H. G., M. H. Oh, B. S. Kim, M. Y. Lee, H. J. Han and S. H. Choi (2009). "The Capability of Catabolic Utilization of N-Acetylneuraminic Acid, a Sialic Acid, Is Essential for Vibrio vulnificus Pathogenesis." Infection and Immunity **77**(8): 3209-3217.
Joerger, A. C., S. Mayer and A. R. Fersht (2003). "Mimicking natural evolution in vitro: An N-acetylneuraminate lyase mutant with an increased dihydrodipicolinate synthase activity." Proceedings of the National Academy of Sciences of the United States of America **100**(10): 5694-5699.
Keown, J. R., M. D. W. Griffin, H. D. T. Mertens and F. G. Pearce (2013). "Small Oligomers of Ribulose-bisphosphate Carboxylase/Oxygenase (Rubisco) Activase Are Required for Biological Activity." Journal of Biological Chemistry **288**(28): 20607-20615.
Konarev, P. V., V. V. Volkov, A. V. Sokolova, M. H. J. Koch and D. I. Svergun (2003). "PRIMUS: a Windows PC-based system for small-angle scattering data analysis." Journal of Applied Crystallography **36**: 1277-1282.
Kruger, D., R. Schauer and C. Traving (2001). "Characterization and mutagenesis of the recombinant N-acetylneuraminate lyase from Clostridium perfringens - Insights into the reaction mechanism." European Journal of Biochemistry **268**(13): 3831-3839.
Kuhm, A. E., H. J. Knackmuss and A. Stolz (1993). "PURIFICATION AND PROPERTIES OF 2'-HYDROXYBENZALPYRUVATE ALDOLASE FROM A BACTERIUM THAT DEGRADES NAPHTHALENESULFONATES." Journal of Biological Chemistry **268**(13): 9484-9489.
Laber, B., F. X. Gomisruth, M. J. Romao and R. Huber (1992). "ESCHERICHIA-COLI DIHYDRODIPICOLINATE SYNTHASE - IDENTIFICATION OF THE ACTIVE-SITE AND CRYSTALLIZATION." Biochemical Journal **288**: 691-695.

Lang, D., R. Thoma, M. Henn-Sax, R. Sterner and M. Wilmanns (2000). "Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion." Science **289**(5484): 1546-1550.
Lawrence, M. C., J. Barbosa, B. J. Smith, N. E. Hall, P. A. Pilling, H. C. Ooi and S. M. Marcuccio (1997). "Structure and mechanism of a sub-family of enzymes related to N-acetylneuraminate lyase." Journal of Molecular Biology **266**(2): 381-399.
McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read (2007). "Phaser crystallographic software." Journal of Applied Crystallography **40**: 658-674.
Nagano, N., C. A. Orengo and J. M. Thornton (2002). "One fold with many functions: The evolutionary relationships between TIM barrel families based on their sequences, structures and functions." Journal of Molecular Biology **321**(5): 741-765.
Panjikar, S., V. Parthasarathy, V. S. Lamzin, M. S. Weiss and P. A. Tucker (2005). "Auto-Rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment." Acta Crystallographica Section D-Biological Crystallography **61**: 449-457.
Panjikar, S., V. Parthasarathy, V. S. Lamzin, M. S. Weiss and P. A. Tucker (2009). "On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination." Acta Crystallographica Section D-Biological Crystallography **65**: 1089-1097.
Pearce, F. G., R. C. J. Dobson, A. Weber, L. A. Lane, M. G. McCammon, M. A. Squire, M. A. Perugini, G. B. Jameson, C. V. Robinson and J. A. Gerrard (2008). "Mutating the Tight-Dimer Interface of Dihydrodipicolinate Synthase Disrupts the Enzyme Quaternary Structure: Toward a Monomeric Enzyme." Biochemistry **47**(46): 12108-12117.
Petoukhov, M. V., D. Franke, A. V. Shkumatov, G. Tria, A. G. Kikhney, M. Gajda, C. Gorba, H. D. T. Mertens, P. V. Konarev and D. I. Svergun (2012). "New developments in the ATSAS program package for small-angle scattering data analysis." Journal of Applied Crystallography **45**: 342-350.
Putnam, C. D., M. Hammel, G. L. Hura and J. A. Tainer (2007). "X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution." Quarterly Reviews of Biophysics **40**(3): 191-285.
Reboul, C. F., B. T. Porebski, M. D. W. Griffin, R. C. J. Dobson, M. A. Perugini, J. A. Gerrard and A. M. Buckle (2012). "Structural and Dynamic Requirements for Optimal Activity of the Essential Bacterial Enzyme Dihydrodipicolinate Synthase." Plos Computational Biology **8**(6): 12.
Renwick C. J. DOBSON, J. A. G., F. Grant PEARCE (2004). "Dihydrodipicolinate synthase is not inhibited by its substrate, (S)-aspartate β-semialdehyde." Biochemical Journal: 757–762.
Riedel, T. J., L. C. Johnson, J. Knight, R. R. Hantgan, R. P. Holmes and W. T. Lowther (2011). "Structural and Biochemical Studies of Human 4-hydroxy-2-oxoglutarate Aldolase: Implications for Hydroxyproline Metabolism in Primary Hyperoxaluria." Plos One **6**(10): 15.
Robert, X. and P. Gouet (2014). "Deciphering key features in protein structures with the new ENDscript server." Nucleic Acids Research **42**(W1): W320-W324.
Schuck, P. (2000). "Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling." Biophysical Journal **78**(3): 1606-1619.
Sello, G. and P. Di Gennaro (2013). "Aldol Reactions of the trans-o-Hydroxybenzylidenepyruvate Hydratase-Aldolase (tHBP-HA) from Pseudomonas fluorescens N3." Applied Biochemistry and Biotechnology **170**(7): 1702-1712.
Sheldrick, G. M. (2010). "Experimental phasing with SHELXC/D/E: combining chain tracing with density modification." Acta Crystallographica Section D-Biological Crystallography **66**: 479-485.
Shibasaki, T., H. Mori, S. Chiba and A. Ozaki (1999). "Microbial proline 4-hydroxylase screening and feme cloning." Applied and Environmental Microbiology **65**(9): 4028-4031.
Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Z. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson and D. G. Higgins (2011). "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega." Molecular Systems Biology **7**: 6.

Svergun, D., C. Barberato and M. H. J. Koch (1995). "CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates." Journal of Applied Crystallography **28**: 768-773.
T.M. Laue, B. D. S., T.M. Ridgeway and S.L. Pelletier (1992). "Analyticall Ultracentrifugation in Biochemistry and Polymer Science." Royal Society of Chemistry: 90-125.
Uchida, Y., Y. Tsukada and T. Sugimori (1984). "PURIFICATION AND PROPERTIES OF N-ACETYLNEURAMINATE LYASE FROM ESCHERICHIA-COLI." Journal of Biochemistry **96**(2): 507-522.
Velasco, A. M., J. I. Leguina and A. Lazcano (2002). "Molecular evolution of the lysine biosynthetic pathways." Journal of Molecular Evolution **55**(4): 445-459.
Vimr, E. R., K. A. Kalivoda, E. L. Deszo and S. M. Steenbergen (2004). "Diversity of microbial sialic acid metabolism." Microbiology and Molecular Biology Reviews **68**(1): 132-+.
Viola, R. E. (2001). "The central enzymes of the aspartate family of amino acid biosynthesis." Accounts of Chemical Research **34**(5): 339-349.
Watanabe, S., T. Kodaki and K. Makino (2006). "A novel alpha-ketoglutaric semialdehyde dehydrogenase Evolutionary insight into an alternative pathway of bacterial L-arabinose metabolism." Journal of Biological Chemistry **281**(39): 28876-28888.
Watanabe, S., D. Morimoto, F. Fukumori, H. Shinomiya, H. Nishiwaki, M. Kawano-Kawada, Y. Sasai, Y. Tozawa and Y. Watanabe (2012). "Identification and Characterization of D-Hydroxyproline Dehydrogenase and Delta(1)-Pyrroline-4-hydroxy-2-carboxylate Deaminase Involved in Novel L-Hydroxyproline Metabolism of Bacteria METABOLIC CONVERGENT EVOLUTION." Journal of Biological Chemistry **287**(39): 32674-32688.
Watanabe, S., M. Yamada, I. Ohtsu and K. Makino (2007). "alpha-Ketoglutaric semialclehyde dehydrogenase isozymes involved in metabolic pathways of D-glucarate, D-galactarate, and hydroxy-L-proline - Molecular and metabolic convergent evolution." Journal of Biological Chemistry **282**(9): 6685-6695.
Wu, H., J. W. Lustbader, Y. Liu, R. E. Canfield and W. A. Hendrickson (1994). "STRUCTURE OF HUMAN CHORIONIC-GONADOTROPIN AT 2.6-ANGSTROM RESOLUTION FROM MAD ANALYSIS OF THE SELENOMETHIONYL PROTEIN." Structure **2**(6): 545-558.